UKRISS

Feasibility Study Final Draft 11th December

December 2012

Author(s)

Richard Gartner, Kings College London
Keith Jefferey, euroCRIS
Richard Jones, Cottage Labs
Lorna Mitchell, Brunel University
Rosa Scoble, Brunel University
Allan Sudlow, British Library
Steve Trowell, University of Exeter
Simon Waddington, Kings College London
Karen Walshe, British Library

1. Introduction

1.1 Purpose and scope of the report

This document is the formal output of the first phase of a study commissioned by JISC to investigate reporting of research information across the UK HE sector and assess the feasibility of any national infrastructure. The scope of the document is to present the outcomes of a nine month investigation into the current landscape of research information management and present options and recommendations for Phase 2 of the UKRISS project that will see the development of feasible solutions for sector-wide research management and reporting. A Glossary of terms used throughout this document is available in **Section 8**.

1.2 CERIF

The Common European Research Information Format (CERIF),[1] developed with the support of the European Commission (EC), is a metadata standard for representation of research information. Technically it is a recommendation by the European Union to its member states. CERIF has been under development since the early 1990s, and has been progressively extended and refined. Since 2002, care and custody of CERIF has been handled by euroCRIS[2], a not-for-profit organisation.

The essential features of CERIF are: (a) it has the concept of objects or entities with attributes such as project, person, organisational unit; (b) it supports n:m relationships between them (and recursively on any of them) using 'linking relations' thus providing rich semantics including roles and time; (c) it is fully internationalised in language and character set; (d) it is extensible without prejudicing the core data model.

Use of CERIF was part of the UKRISS terms of reference specified by JISC.

1.3 Background

The JISC-funded EXRI-UK[3] study of 2009 conducted a review of available standards for representation and exchange of research information. It recommended the adoption and further development of CERIF (the latest version was CERIF 2008) in the UK HE sector. The recommendations included developing pilots to demonstrate the application of CERIF in specific use cases.

The EXRI-UK study was supported by a further study commissioned by JISC in 2010 to examine the business case for CERIF adoption[4]. It concluded that the overall cost of either deploying a CERIF-compliant CRIS or writing CERIF wrappers around non-CERIF compliant institutional and funder systems was low in relation to the benefits that could be realised in terms of reduced complexity of information exchange, compared to exchanges in multiple ad hoc formats.

---

[1] CERIF: http://www.eurocris.org/Index.php?page=CERIFreleases&t=1#
[2] euroCRIS: http://www.eurocris.org
[3] EXRI-UK Report: http://repository.jisc.ac.uk/448/
[4] JISC Bolton report (2010): http://www.jisc.ac.uk/publications/reports/2010/businesscasefinalreport.aspx

A further JISC-funded report[5] in 2012 examined the adoption (by early 2012) of CERIF-compliant systems within UK HE institutions. At that point, adoption was around 30% of UK HEIs. All but one of these CRIS systems was supplied by a commercial vendor. The three CERIF-compliant commercial systems deployed were PURE (Atira)[6], Elements (Symplectic)[7] and Converis (Avedas)[8]. Despite the long standing availability of CERIF as an underlying standard, many institutions have not engaged with CERIF directly, due to the complexity of the standard and lack of time and resources.

In parallel to the deployment of institutional systems, funders have been developing systems to capture research outputs. Five of the RCUK funding councils use the Research Outcomes System (ROS)[9], which was developed in-house and is implementing CERIF wrappers for exchange of data. Work is on-going to align the information collected by each of the councils using ROS. MRC, STFC and also Wellcome Trust have implemented a different outputs system run by a commercial vendor (Research Fish[10]).

The Gateway to Research[11] (GtR) project, funded by BIS, aims to develop a public portal that collates a subset of the research information from the research council systems into a single repository to provide a showcase for UK research outputs. GtR is performing mappings to CERIF as part of the project.

Following the EXRI-UK recommendations, a number of JISC projects have developed prototypes for specific use cases using CERIF including Readiness 4 REF (REF submissions), BRUCE (report generation and benchmarking), CERIF in Action (upload to ROS and institution-to-institution information transfers) IRIOS1 and 2 (linking outputs to grants) and RMAS (information exchanges between internal institutional systems).

The 2012 HESA-funded Information Landscape Study[12] looked at a wide range of information requests made to institutions, particularly focusing on students, to identify ways of reducing the data collection burden. The report evaluated several options for simplifying information requests, including use of a single information collection agency. The overall conclusion was that the solution should harness the collaborative culture that already exists to improve efficiency, rather than imposing a centralised governance model.

1.4 UKRISS project

UKRISS is a JISC-funded project, which is conducting a feasibility and scoping study for the reporting of research information at a national level based on CERIF, with the objective of increasing efficiency, productivity and quality across the sector. Our aim is to define and prototype solutions which will be compelling, easy to use, have a low entry barrier, and support innovative information sharing and benchmarking.

---

[5] JISC Rosemary Russell report: http://www.ukoln.ac.uk/isc/reports/cerif-landscape-study-2012/CERIF-UK-landscape-report-v1.1.pdf
[6] Pure: http://atira.dk/en/pure/
[7] Elements: http://www.symplectic.co.uk/
[8] Converis: http://www.avedas.com/
[9] Research Outcomes System: http://www.rcuk.ac.uk/research/Pages/ResearchOutcomesProject.aspx
[10] Research Fish: http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/e-Val/index.htm
[11] Gateway to Research project: http://www.rcuk.ac.uk/research/Pages/gtr.aspx
[12] HESA Information Landscape Study: : http://landscape.hesa.ac.uk/wp-content/uploads/2012/01/IRPG_PrjB_Final_Report.pdf

The scope of potential solutions to be investigated includes both the development of national-level services and infrastructure as well as solutions not requiring infrastructure development such as harmonisation of data formats and dictionaries, which meet the overall aims of achieving efficiency, productivity and quality gains.

The project runs from 1st March 2012 to 30th June 2013. This report is the main deliverable from Phase 1, which runs until 30th November 2012. If the recommendations are accepted, Phase 2 will see the development of applicable solutions that will benefit, both in terms of efficiency and effectiveness, research management and reporting across the sector. Further detailed information about the project lifecycle and deliverables can be found in **Appendix A**.

1.5 Document outline

The document consists of a number of shorter main sections with further detailed information and source materials placed in Appendices.

The overall structure of the document is as follows. **Sections 2** to **4** describe the work undertaken in Phase 1, while **Sections 5** and **6** propose different options that have emerged from Phase 1 and recommendations for Phase 2 of the project. Details of sections are:

From Phase 1: **Section 2** contains a description of the key stakeholders, reporting landscape and technology landscape. This work was carried out at an early stage in the project to obtain an overview of the current status of reporting and identify areas for further investigation. The project conducted an extensive user requirements gathering study. The methodology of the study is described in **Section 3. Section 4** identifies the drivers, use cases and requirements extracted from analysis of the raw data collected in the study.

For Phase 2: **Section 5** describes the individual tasks or components that have been defined to meet the requirements extracted during the UKRISS study. **Section 6** describes how these components can be combined into alternative work plans and presents a business case for the main areas of work. **Section 7** presents the conclusions and recommendations for Phase 2, as well as comments and recommendations from the UKRISS Steering Board.

2. Research information landscape

This initial phase was the basis to plan and coordinate the capture and analysis of requirements for the delivery of Phase 1. Two landscape studies were conducted to give a full picture of current CRIS systems, projects and developments both at a national and international level.

2.1 Stakeholder analysis

A stakeholder analysis was conducted in parallel to the landscape study outline below. The purpose of the stakeholder analysis was to identify the key players who were influencing and/or were influenced by developments in the research reporting infrastructure in the UK.

The methodology for the identification and analysis of stakeholders is described in **Section 3.1**. Thus identified, key stakeholders were approached for interview to gather requirements and drivers as described in **Sections 3.2** and **3.3**.

Appendix B describes how stakeholders were categorised into a number of typologies and analysed for predicted interest in, and potential impact on, the recommendations arising from the project. Aside from 'Project Sponsor' i.e. JISC, typologies comprised the following:

- Funder - Government
- Funder - Charity
- Umbrella Organisation
- HE Institution
- CRIS Vendor

Summary descriptions on these typologies are provided in **Appendix B**.

The project team recognises that these typologies are a blunt instrument i.e. an individual organisation will have influences and interests that extend beyond these categorisations. It was also recognised that time constraints would not permit the project to consider all the stakeholders that are players in the reporting landscape (see below) in detail. Thus, it was decided to focus on research reporting in the context of HE institutions and funder relationships with one another, with CRIS vendors, and with (umbrella organisations) with an overarching interest in improving efficiency in this domain.

The overarching drivers that emerged from stakeholder consultation and analysis are referenced in the table in **Appendix B** and discussed in detail in **Section 4.3**.

2.2 Reporting landscape

The landscape study aimed to provide an overview of current and developing projects and systems for the management and reporting of research information and the relationships between these. This included the systems and processes involved in the current statutory returns as well as institutional (i.e. internal and local systems), national and international CRIS systems and repositories.

The landscape study was used in association with the stakeholder analysis and requirements gathering exercise to ensure that the project's recommendations were directly relevant to the environment in which researchers and administrators are actually working. Full details can be found in **Appendix C.**

A long list of 60 systems was identified and a taxonomy developed to target projects/systems that were either statutory in nature or had key characteristics common across the sector. This was to ensure that any future requirements gathering did not overlook the statutory need of funders and essential features already embedded at sector level. For example:

- Requirements for information from funders such as the one that has to be provided through Research Outcomes System (ROS) on outcomes of RCUK funded research
- Systems that already provide information sector-wide on funded projects such as Grants on the Web[13]
- Development such as VIVO[14] an open source semantic web application that can be populated with researcher interests, activities, etc. and enables the discovery of research and scholarship across disciplines and institutions.

The landscape study also mapped some of the international systems, projects and developments. While this section could not be explored in detail because of time constraints some information on international approaches were collected. For example, it was reported that the Czech Republic deploys a centralised government management system where funders report annually on all their funding activities via a universal service[15]. The central database comprises a registry of funded projects, institutional research plans, research outputs (publications, patents and technologies) and funding calls. Analytical tools are provided for use by government officials and funders. The largest section is research outputs with 870,000 objects, increasing at a rate of 100,000 per year. Apart from certain classified data and personal information, all material is publicly available on the web. Bulk export in zip format is possible, and search capability is provided.

The outcome of the study was to ensure the scrutiny and consideration of systems and projects:

- Aimed at embedding CERIF as a research information exchange standard; mainly JISC funded projects
- That presented sector wide solutions for research information management and the harmonisation of research data
- For the collection of information for statutory returns
- That could present international best practice.

It should be noted that whilst a minority of institutions are using dedicated systems to manage research information (c.f. **Section 1.3**); the norm is still for this information to be stored within other systems, e.g. HR databases, finance systems, student information systems, etc. It is also worth noting that research

[13] Grants on the Web: http://gow.epsrc.ac.uk/

[14] VIVO: http://VIVOweb.org/

[15] Chudlarský, Tomáš; Dvořák, Jan: *A National CRIS Infrastructure as the Cornerstone of Transparency in the Research Domain.* In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 9-17. ISBN 978-80-86742-33-5. http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_1_full_paper.pdf

information management systems are not currently viewed as being “key” within HEIs and institutions are therefore less likely to be willing to commit funds to developing them.

2.3 Technology landscape

As part of a more detailed study into the landscape, we conducted in-depth technical reviews of a number of key technologies in different aspects of research information management: CRIS systems, funder systems, business intelligence, information modelling, and related JISC-funded projects. In all we investigated 12 technologies/projects; more complete details of which can be found in **Appendix D**:

- Atira Pure - Institutional CRIS system
- Avedas Converis - Institutional CRIS system
- BRUCE/SolrEyes[16] - JISC funded research information project
- CERIF - Research information data model
- InCites[17] - Research analytics service
- Je-S[18] - Research Council grant submissions system
- MICE[19] - JISC funded research information project
- Readiness 4 REF[20] - JISC funded research information project
- Research Fish/e-Val - Funder CRIS system
- RMAS[21] - JISC funded research information project
- ROS (Research Outcomes System) - Funder CRIS system
- Symplectic Elements - Institutional publications management system

The aim of this work was to survey the kinds of functionality that we find in well-developed widely used systems, and what potential implementation options we might have for any Phase 2 development work. What we found was an extensive feature set which includes: authority lists for authors (on an institutional level), structured databases of citations and publications and flexible reporting and management information tools. This information will be useful later as we look to make recommendations in the second part of the project.

2.4 Commentary

Figure 2.1 below illustrates the findings of the project’s landscape studies and demonstrates the complex nature of the current research information management processes within the sector. The landscape shows interaction between research activities and information between and within three broad environments: researcher, institutional and external. At researcher level, most activities need to be reported many times to both the institution and externally; for example Knowledge Transfer (KT) activities need to inform both the institution and also be reported in HESA’s Higher Education Business and Community Interaction (HE-BCI)[22] survey, and Postgraduate Research (PGR) activity has to be captured at many levels. There are also instances where single items need to be captured by multiple systems such as outputs that might need to be entered in institutional CRIS systems, in funders systems

[16] BRUCE: http://bruceatbrunel.wordpress.com/
[17] InCites: http://researchanalytics.thomsonreuters.com/incites/
[18] Je-S: https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx
[19] MICE: http://mice.cerch.kcl.ac.uk/
[20] Readiness4Ref: http://www.kcl.ac.uk/innovation/groups/cerch/research/projects/completed/r4r.aspx
[21] RMAS: http://www.exeter.ac.uk/research/rmas/
[22] HE-BCI: http://www.hefce.ac.uk/whatwedo/kes/measureke/hebci/

(e-Val, ROS, etc.), institutional repositories, open repositories, Research Excellence Framework (REF) submission system, etc.

There is currently very little effective sharing of this information with the result that the same information may be sent to multiple systems in several different formats, generating work and the potential for errors to be created.

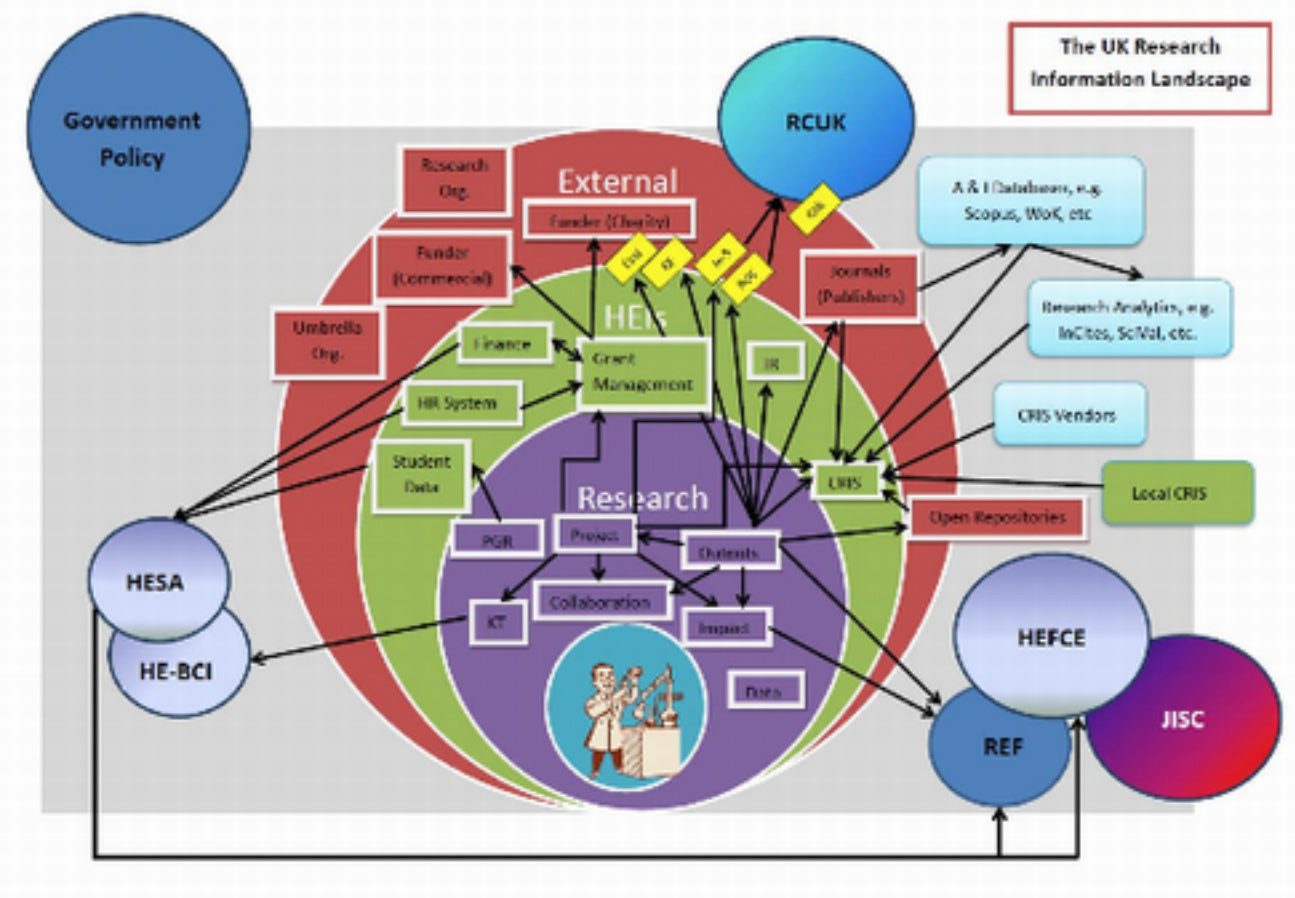

Figure 2.1: Landscape study findings

3. Identification of Stakeholders, Drivers and Requirements

This section outlines the methodology that underpinned identification of stakeholders, and the capture and analysis of their drivers and requirements. The deliverables at each stage are also highlighted.

3.1 Stakeholder analysis

Deliverables

- A stakeholder matrix of relevant organisations, professional associations, umbrella bodies, along with their interest, potential impact and consultation routes in relation to the project – **Appendix B**
- A list of individual named stakeholders to be approached for interview.

Methodology

Informed by the landscape study, the project team generated a long list of relevant individuals, Research Organisations, funding bodies and professional associations representing key stakeholders in the domain. The emphasis for scoping down this long list to a feasible number of stakeholders to approach was based on who had deployed, funded or was using a system to support research information management.

Thus identified, stakeholders were categorised into broad typologies based on role type in relation to the project. Stakeholders were then stratified within these broad typologies to ensure a representative sample of organisations across sectors of different sizes and maturity.

As a final step, the project team consulted with the project steering board to validate this stakeholder analysis and thus confirm the identity of interviewees for the next stage of the consultation.

3.2 Interview design

Deliverables

- A comprehensive set of 64 interview questions, mapped to typologies, as well as functional and non-functional requirements, to maximise the utility of the qualitative information captured for translation into requirements across different types of stakeholder - **Appendix E**.

Methodology

With advice from the steering board, the project team developed a long list of questions that mapped to typologies to support analysis across stakeholders to look for commonalities, differences and trends.

These questions were tailored during interview to fit the roles and responsibilities of interviewees. The question set also aimed to map to a broad set of functional and non-functional categories that would impact on service/solution development, adoption and sustainability. To help identify the current status of the landscape in more detail, the question set also covered the scale, scope and (non-) maturity of the organisation's research information management systems.

After refining, the question set was checked for topic gaps, duplication or redundancy. The finalised questions were then rechecked to ensure they were in scope and mapped to one or more of the requirement categories. The project steering board was then provided with the question set for review/ approval.

3.3 Interview process

Deliverables

- A set of interview recordings and transcriptions
- A set of cross-referenced raw requirements, drivers and technical implementation notes from each interview.

Methodology

Face-to-face structured interviews were the default capture methodology, with telephone interviews as a standby option. This approach was chosen in preference to others, such as a self-completion survey, to enable the project team to tailor interviews to fit the roles and responsibilities of interviewees. In person interviews also allowed the project team to capture tacit knowledge more effectively, probe on previously unidentified areas that became apparent during the interview and build rapport with the interviewee.

Initial approach to interviewees was via a personal email, tailored to help elicit a favourable response, which included an outline of project aims, a handful of priority topics to prepare the recipient in advance of interview and notification of the likely duration and format of the interview.

A consent form setting out the details of intended use of interview material and interviewee's consent on various matters was also sent out in advance – **Appendix F.**

Responsibilities for administering each stakeholder interview were agreed in advance across the project team, taking account of domain knowledge and personal contacts.

Face-to-face interviews were conducted at interviewees' workplaces which lent context to discussions. Interviews lasted – on average – for 1 hour, and were audio recorded. Transcripts were produced from these recordings via an approved professional agency and quality assurance carried out by the project team. The interviews generated around 900 pages of transcribed text which went through to analysis.

Post-interview personal email provided an opportunity for follow-up and/or clarifications and included an offer to share the transcription with the interviewee, a copy of the interview analysis and highlights of points that the interviewee may have considered important.

3.4 Interview analysis

Deliverables

- A series of spreadsheets – one per interview – containing references to statements within the interview transcripts coded against requirements, drivers and technical implementation notes
- A meta-analysis across the aggregated interviews to identify and summarise the key drivers and requirements – **Appendices G** and **H**
- A list of high-level requirements and key use cases listed in **Section 4**. These are cross referenced to the full set of requirements in **Appendix K**
- A summary of the overall analysis set out in this document.

Methodology

Transcripts were analysed line-by-line, with statements being coded against requirement and driver categories as outlined below. Raw requirements extracted from each interview were clustered according to a two level semantic hierarchy contained in **Appendix I.**

All the source requirements were linked to a page in the interview text, so all the requirements are traceable to their original sources. The requirements were de-duplicated within each category. De-duplication was not carried out across categories, since source requirements could have multiple contexts. The requirements were linked to the typologies of the stakeholders that generated them. Thus, for example, it is possible to check that a given requirement was requested by three organisations, two of which were Russell group members. The final requirements generated are contained in **Appendix H**. Finally, a list of eight high-level requirements were generated together with a summary of the key use cases. The cross references from the high-level to the source requirements is described in **Appendix K.** The eight main requirements are summarised in **Section 4.4.**

A set of 209 driver statements – motivations underlying the interview responses – were identified during the interview analyses. Driver statements were coded according to the following categories:

Strategic	political, competitive, marketing, research drivers for business delivery
Operational	day-to-day practical, workflow, management, implementation, efficiency drivers
Technical	technology, functional, hardware, software, standards drivers
Economic	cost, saving, resourcing drivers
Governance	statutory, legal, ethical, contractual drivers
Reporting	reporting, submission and similar transactional drivers
Social	user, engagement, adoption, transition drivers

Table 3.1: Driver statement categories

Drivers were filtered according to categories, analysed for common themes, de-duplicated and coalesced around a small number of overarching drivers sharing a common description format. **Section 4.3** provides a summary of these overarching drivers, as well as an analysis on the different stakeholder perspectives in relation to these drivers.

4. Use cases and business needs

Sections 4.1 and **4.2** describe the use cases that were uncovered during the requirements gathering process that are dependent on exchanges of research information between institutions, funders, charities and government bodies. The use cases are described from the perspective of both institutions and funders (of all types). Each use case includes a brief description, the required data exchanges and how the use case is relevant to UKRISS.

Section 4.3 presents the stakeholder and business driver analysis derived from the UKRISS study. **Section 4.4** contains a summary of the main requirements from the study. Further detailed information is referenced in the appendices.

4.1 Institutional use cases

This sub-section describes use cases relating to data transfers to and from institutions, and describes any value adds that could be made to enhance that use case.

Reporting to Funders: In order to report to funders, information has to be collated from multiple institutional systems (finance, HR, CRIS, repository) and a bulk upload must be done to the funder's system. The data can be complicated, and can include reporting on co-funded work. It would be of value for this use case to simplify the gathering of information from internal systems, and reduce the duplicate effort required to report to multiple funders, while increasing automation. Institutions are also interested in data validation prior to submission to the funder.

REF reporting: REF is a critical but infrequent activity for institutions. REF requires institutions to interact directly with HEFCE. It would be valuable to have a standard format in which the reporting could be done, and in which the reporting requirements were harmonised with other funder reporting activities.

Statutory reporting to HESA: Institutions have to report regularly to HESA focussing on elements relating to research staff and students. It would be valuable if HESA reporting were in line with the reporting requirements of other external bodies (such as funders) and processes (such as REF) to reduce duplication of effort.

Internal reporting, benchmarking and operational management (e.g. optimising facilities usage) : Management information requires reports to be generated based on research information and other operational data collected from within the organisation. This requires many different types of system to talk to each other (e.g. CRIS, HR, finance), which indicates the need for alignment of representation of internal data, and possibly the support of benchmarking tools.

Benchmarking against other institutions: Institutions want to be able to compare research information collected internally with information acquired from funders or directly from other institutions. There would be value in a core information profile across which comparisons could be made, and some standard metrics which could be extracted from the comparison. There is also a need for a tool-chain which can consume data in appropriate formats (e.g. CERIF) and provide an interface over the top.

Portfolio analysis (internal) and collaboration (cross-institution): It is useful for management information purposes to be able to analyse an institution's current research activities, and from that to identify opportunities to collaborate with other institutions. This requires data to be grouped internally (by department, research field or collaborative network), and then compared/contrasted with data obtained from other institutions.

Submission of grant proposals to funders: Institutions need to submit grant proposals to funders, and there are a variety of ways of doing this, often involving duplication of effort and re-keying information. It would be useful, for example, to be able to connect institutional systems directly to the J-eS submission system.

Researcher CV generation: Institutions, funders,and researchers themselves want to be able to generate a researcher CV based on their full working life (not just their employment within a single institution), which implies an exchange of data across institutional boundaries when researchers move, and a Researcher ID of some form (e.g. ORCID). There would be a lot of value in automating this process, and for the CVs themselves to be auto generated (which is already partially the case in many institutions). This data may then be submitted to a funder along with a grant application, so being able to automatically convert and submit the CV in the appropriate format would be useful.

External communications: Institutions want to be able to make publicly available a subset of their research information, typically via a website or an API. This raises issues of quality control and authority, as data travels out into the open, beyond the institutional firewall, and the workflows required to transform data into this state.

Staff performance management: Institutions need to collect and analyse research information pertinent to staff appraisals, which can be used either to inform staff development or promotion panels. Much of the data is available internally, but there could be value in acquiring some of this information from funders.

Strategic planning (future): Based on existing research information, institutions need to carry out analyses which allow them to plan for the future (e.g. identify new research opportunities). This means gathering information from a variety of sources including funders, charities, and commercial data sources. Harmonised data formats could assist in analysing across multiple data-sources and disciplines; the data needs to be timely and of high quality.

Compliance monitoring: In order to ensure they qualify for future grants, institutions are concerned with monitoring reporting compliance, delivery of work, and financial targets. Improved communication between institution and funder could make it easier to be compliant and identify compliance issues, while automated aggregation of data from institutional systems could make the compliance monitoring easier.

4.2 Funder use cases

This section covers use cases relating to funders of all types (from government-backed to charities), that have the capability to accept research information from institutions, and describes any value adds that could be made to enhance that use case. Use cases involving funders that also involve institutions are covered in the previous section and not included here.

Publication of source data via common APIs: Funders want to be able to clean and publish data collected from institutions, so that it is available via APIs (including making it available to Gateway to Research).

Funder-funder benchmarking: Funders wish to be able to compare their relative performance, and potentially HEFCE/BIS will want to do the same. This requires data to be available across funders, to enable benchmarking (across a core set of comparable data) with some defined metrics. Availability of benchmarking tools to aid this would be valuable.

Portfolio analysis and strategic planning: Funders want to be able to analyse their current research activities, and identify opportunities for future funding calls. This mostly requires funder-internal data (or that gathered from institutions), but also potentially data from other funders. Having access to such data in a harmonised format would therefore make these tasks much easier.

Reporting to government (regular and ad hoc): Funders are often required to report upwards to government, and are reporting on data which in turn has been reported to them via institutions. Access to timely and high quality data is essential to enable rapid and reliable report generation.

External communications: Funders want to be able to make publicly available a subset of research information, typically via a website or an API. This raises issues of quality control and authority, as data travels out into the open, beyond the funder's firewall, and the workflows required to transform data into this state.

Gathering of information from researchers: Researchers need to be able to directly enter research information via a web form or some other manual means. This data should be harmonised with the kind of data that is uploaded by researchers in bulk uploads, and should have similar quality control measures in place.

Production of evidence of impact: Researchers need to be able to directly enter textual information describing the wider impact of research.

4.3 Stakeholder and driver analysis

The complexity of the current research reporting landscape was outlined in the previous section. In this section we examine the stakeholders who influence this landscape and focus on their drivers and impacts.

Stakeholders for the project were identified and categorised into broad typologies according to the methodology outlined in **Section 3.** The stakeholders we interviewed were chosen to represent a range of relevant strategic, policy/operational and technical perspectives on research information management. The stakeholder matrix, which details the HE institutions, Research Organisations, funding bodies and commercial vendors we interviewed can be found in **Appendices B and I.**

4.3.1 Overview summary

Our study revealed a research information management landscape in the UK that is currently fragmented. Intentions did not match realities. Stakeholders from across the sector that we interviewed – be it research funder, government agency, HE institution or software vendor – were, with good intention,

implementing systems and processes to support their research information needs as best they could. All were aiming towards the same goal of improving the efficiency and quality of information management and reporting to enable more agile, evidence-based decision-making. However, because of differences in motivations for designing and implementing these systems and processes, a joined-up approach within and across organisations was often lacking. While perspectives and drivers differed across stakeholder groups, a number of shared themes, described below, emerged.

4.3.2 Harmonisation

Consortia from the different stakeholder groups had begun the drive towards harmonisation to help reduce the reporting burden and enable cross-sector impact analysis and evaluation. There was a shared awareness amongst funders of the need to articulate the return on investment from the public purse through better evidence gathering and reporting. Harmonisation was seen as a driver to support this.

There was an inherent tension in this desire to harmonise. Funders and HE institutions wished to benchmark themselves against others but also wanted to converge on reporting standards and processes that represented them in the best light. An additional consideration was the level of granularity at which they were able to share information and data. This arose from a range of sensitivities relating to the corporate, commercial or personal confidentiality associated with some of their information and data assets. There was also a transition challenge to harmonisation as existing funder and HE institutions systems were at different levels of maturity and adoption. Thus, adapting already tailored systems and processes to enable interoperation in the current landscape is as much a social as a technical challenge.

4.3.3 Costs

Prior to the stakeholder interviews, we had surmised that a major driver for greater harmonisation would be economic. The shared goal of reducing the reporting burden on researchers and research administration by greater automation of information management was recognised as important but not described by stakeholders in the context of cost savings. Stakeholders were more interested in efficiency gains. However, on-going costs for the sustainability of such solutions were a consideration. Some stakeholders wanted their in-house systems to capture and track information on published outputs more efficiently to remove the need to purchase commercially-sourced bibliometric data.

4.3.4 Research quality

Many stakeholders, particularly those responsible for overseeing research in HE institutions, indicated that improving the quality and impact of their institution's research was their key driver. All stakeholders recognised that better quality research information was essential to enabling this improvement. This research information underpinned their business intelligence, and its quality, presence or absence had a significant effect on their ability to plan and manage their research portfolio. Good business intelligence also allowed stakeholders to demonstrate value, exploit strategic gaps and opportunities, and remain competitive.

4.3.5 Solutions and standards

It was clear that whilst HE institutions wanted to deploy solutions that helped them to deliver statutory and funder reporting as efficiently as possible. They also wanted these solutions to enable them to capture quality-assured information and re-use it in many different research strategy and planning contexts. The ease of integration, scalability and flexibility to business needs were important considerations for solution acquisition and development. Information standards which might enable this were mentioned by those

involved in deployment and management of research information systems, but CERIF had low visibility amongst this stakeholder group. CERIF compliance was not a driver for funder or HE institutions choice of systems.

4.3.6 Information flow

Information flow, or rather lack of it, was a recurrent theme. Many of those consulted wanted to improve information flow so that it became a two- rather than a one-way process, for both internal and external systems. Funders and research management teams in HE institutions saw this enabling richer, more agile reporting and analysis. Good information flow was seen as critical to reducing reporting burdens, enabling feedback and supporting monitoring. In contrast, the researchers that we interviewed did not perceive any reduction in the requests they were receiving currently to provide information to colleagues or enter it themselves online.

4.3.7 User adoption

Motivations for researchers to upload their information were mostly compliance-based i.e. they did it because their funder or institution required it. An emerging driver in this regard was observed where an institution's centralised system was the only place where information was sourced for performance reviews, promotion panels, REF submissions and other reporting that impacted on an individual researcher's career. Benefits-based drivers for adoption included auto-generation of CVs and web profiles that were configurable, and the more general ability for researchers to extract and re-use the information they had submitted. Researcher adoption was closely linked to the ease of use of the system interface.

4.3.8 Analysis

The methodology for capture and analysis of the stakeholder drivers is set out in **Sections 3.3** and **3.4**. A spreadsheet representing a meta-analysis of aggregated drivers across all interviews is provided in **Appendix G**. Below we outline the outcomes of this analysis in more detail.

A set of 209 driver statements – motivations underlying the interview responses – were identified during the analyses. Driver statements were coded against one or more categories, and their frequency of occurrence recorded thus:

Driver Category	Definition and Scope	Occurrences
Strategic	political, competitive, marketing, research drivers for business delivery	77
Operational	day-to-day practical, workflow, management, implementation, efficiency drivers	133
Technical	technology, functional, hardware, software, standards drivers	57
Economic	cost, saving, resourcing drivers	28
Governance	statutory, legal, ethical, contractual drivers	26
Reporting	reporting, submission and similar transactional drivers	60
Social	user, engagement, adoption, transition drivers	44

Table 4.1: Frequency of occurrence of driver statements

Drivers were filtered according to these categories, analysed for common themes, de-duplicated and coalesced around six overarching driver descriptions sharing a common format:

[overall aim] *through* **[improvement]**

Thus:

D001 Improve business intelligence, management and due diligence *through* better information quality and reporting utility

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community *through* harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development *through* systemisation and harmonisation of reporting

D004 Increase research community reporting compliance *through* deploying easy-to-use, flexible reporting systems with user benefits

D005 Improve the research, strategy and planning across UK institutions *through* use of better quality reporting information

D006 Improve research information management across the sector *through* deploying sustainable, affordable solutions that are fit-for-purpose

Each of these broad statements reflects a range of motivations from across the stakeholder groups. Below we draw out these motivations in more detail.

4.3.9 Themes and examples

D001 Improve business intelligence, management and due diligence *through* better information quality and reporting utility

For both funders and HE institutions, improving business intelligence means better research portfolio management to inform strategy and planning. In effect, knowing at any point in time, what research they are managing, why, how much and what it has/will deliver. Business intelligence is also about these stakeholders being able to capture and validate the whole range of activities, outputs and impacts of an individual researcher to enable informed decision-making and performance management. Reporting utility for research institutions goes beyond the REF and refers to an ability to collect information once and repurpose it for a wide range of different internal and external reporting requirements. Possessing good business intelligence allows more effective communication of outputs and impacts to a wide range of stakeholders, including the general public. It enables easier compliance for due diligence purposes such as reporting use of research funds, IP and contract management, and responding to FOI requests. All these aims were seen to be dependent on improving research information quality.

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community *through* harmonisation of reporting processes and/or systems

Reducing the reporting burden on the research community was seen as an efficiency driver: reducing the administrative costs and effort and allowing researchers to spend more time engaged with research. Efficiency was also one of the drivers for cross-funder harmonisation. Multiple stakeholders saw value in

greater consistency for research information reporting - systems, software, standards – and interoperation with internal and external systems to reduce the need for manual effort.

HE institutions in particular recognised that standardisation was not enough. Having a common language for describing the research information being collected, managed and shared was also seen as key to reducing the reporting burden. This common language (e.g. an agreed definition of 1 FTE research) needed to be easily digestible and acceptable to those undertaking the collection and management, as well as the evaluation of the research. Aligned to this, HE institutions were keen to see funders agree a standard core set of reporting requirements across the board; it was recognised that this would be supplemented by a smaller level of funder-specific reporting requirements.

D003 Enable cross-sector impact analysis, evaluation and strategy development *through* systemisation and harmonisation of reporting

In addition to the efficiency gains resulting from systemisation and harmonisation as outlined above, other benefits were identified as drivers for change. Enabling cross-sector analysis was one of these. For funders, this helped them co-ordinate investment across the research landscape to maximise their impact and strategic positioning. Research institutions also identified these benefits but in the context of remaining competitive. Benchmarking was recognised as a specific approach that would be easier to undertake with greater harmonisation. Consistency in the interpretation of the research information that was collected was also seen as important. This was linked to agreeing a common framework for defining and evaluating research outputs and impacts.

D004 Increase research community reporting compliance *through* deploying easy-to-use, flexible reporting systems with user benefits

A number of compliance-based motivations and functionality benefits for users that promote adoption of reporting systems have already been summarised above. For reporting systems reliant on researchers or research office staff inputting information, ease-of-use remained a key driver for adoption. System flexibility was also important as users wanted to tailor systems to different institution needs and run different types of reports at different points in time. Convergence of systems, to the extent of having a single interface for multiple funders reporting needs, cropped up several times. The tension here is that most research institutions and funders are already committed to deploying a range of different reporting systems.

Communication, or rather lack of it, was a barrier to compliance – funders were not always being explicit about how requested information would be used nor communicating back to institutions when reports had been submitted and approved. Better information flow between reporting systems, both in terms of ease of interoperation, and an ability to retrieve information for re-use in different contexts was a positive driver for adoption.

D005 Improve the research, strategy and planning across UK institutions *through* use of better quality reporting information

This driver was closely aligned with the benefits derived from improving business intelligence, some of which have already been summarised above. This included an ability to capture structured research information from different funding streams for monitoring, evaluation, benchmarking, forward planning, policy work and strategy development. Research institutions mentioned a range of research portfolio analyses enabled by better quality reporting information that they perceived as beneficial. PVCs and research offices highlighted a need to identify areas of current and emerging strength to inform development of critical mass and responses to funding opportunities. This was aligned with a general

ability to move research information management away from a sole focus on retrospective reporting towards evidence-based forward planning. Good business information was also seen as an enabler for developing new collaborations, joint activities and networking between organisations and individuals in different research sectors, both on a national and international scale.

D006 Improve research information management across the sector *through* deploying sustainable, affordable solutions that are fit-for-purpose

An overarching driver for deployment was ensuring that the investment in integrating and setting up new research information systems didn't outweigh the benefits. Investment in this sense was more often identified as effort, time and human resource rather than cash expenditure. Another barrier to adoption was a general wariness amongst some institutional IT staff of the business need for deployment of new systems and the consequent need for ongoing support in terms of costs and skills. Other concerns pertained to systems integration and future-proofing to ensure fitness for purpose. The focus in many institutions in the past had been on development of systems to meet transactional administration needs rather than portfolio management and analysis.

Obviously resource constraints across the different research institutions consulted varied in magnitude and nature. These constraints applied to funders as well, with an additional consideration of how they want to manage their relationship with their fundees. For example, smaller charity funders were concerned that a centralised system might present a barrier to the close relationship they currently have with their research community.

4.4 Requirements analysis

In this section, we describe the eight main requirements extracted from the UKRISS study, using the methodology in **Section 3**. Requirements are labelled R001-R008. Each requirement has a high-level description supplemented by a more detailed set of sub-requirements. The derivation of the final requirements from the source requirements is contained in **Appendix K**. A full list of the source requirements from the study is contained in **Appendix H.**

R001 Harmonise dictionaries and usage of CERIF within the UK HE sector

a. Produce a common set of definitions of data dictionaries, output types (including non-publications, identifiers (people, equipment, grants, and funders), institutional structures, research topics and metrics.
b. Specify use of DOIs for linking outputs and equipment to grants and funders, outputs to researchers etc.
c. Align more closely standards development and implementation with the practical requirements of a wide range of stakeholders.
d. Support international initiatives such as ORCID, FundRef[23] and CrossRef[24].

The study uncovered numerous issues regarding the usage and implementation of CERIF within both institutions and funders. CERIF is a powerful and flexible schema for representing research information. It enables a wide range of information to be represented and complex relationships to be modelled. However, the precise mapping of information fields to CERIF entities often resulted in ambiguities, leading to lack of interoperability. There is a reliance on data dictionaries or standard terms. Since these are not provided within the standard itself, there is currently no uniform set of definitions for the UK HE

[23] FundRef: http://www.crossref.org/fundref/index.html
[24] CrossRef: http://www.crossref.org

sector that can be applied.

There was recognition of the need for standard identifiers, both for people as well as other entities such as equipment and grants. There was an awareness of international initiatives such as ORCID and FundRef, and that there is a need for both national and international approaches. Increasing need is being made of DOIs, to enable automated processing and in particular association of research outputs with research grants. Areas such as organisational structures and research topics are recognised as complex areas that require further work. There was a clear wish to extend to range of available output types that can be represented in CERIF, both to include greater expressiveness for certain disciplines as to represent a wider range of non-publication outputs. Not all key stakeholders were directly engaged with CERIF standards development, and there was a clear need to collect a more exhaustive set of requirements that could be incorporated into the standard. There was also recognition that standards development is quite slow and costly, and not all stakeholders had the capacity to engage fully.

R002 Obtain agreement between all key stakeholders (e.g. funders, institutions, charities, statutory bodies) on closer alignment of reporting requirements and their persistence, and adoption

a. Define a minimum core dataset that is collected by all stakeholders to enable comparison, sharing and re-use.
b. Enable reporting information to be collected once and associated to multiple funders.
c. Develop agreed definitions of non-publication outputs and impact measures.
d. Align funder, institutional and charity reporting requirements with those of statutory reporting such as HESA returns and REF.
e. Ensure compliance with agreements to collect a minimum core dataset.

There was strong consensus on the need to provide a common set of information fields that could be collected by all funders. This would simplify the reporting process by reducing the duplicate reporting that is currently required by both researchers and institutions. Much research is now interdisciplinary resulting in an increase in co-funded projects, and there is a need to simplify their reporting. There was also recognition that such a common set of fields would provide a basis for more effective benchmarking across the sector.

There is strong interest in collecting a wider range of non-publication outputs and agreement on common definitions of such outputs is required. In particular, similar outputs are often classified differently across different research disciplines. A standard set of information fields should take into account requirements of funders and charities, institutions and statutory bodies such as HEFCE and HESA.

Work on the 2014 REF is now well advanced. The timing or requirements for future REFs are unknown, so their requirements could not be integrated. However, there was a desire that future REFs should factor harmonisation that is occurring across the sector into their planning. There were differing views on whether a common reporting profile could be achieved purely through consensus, or whether some degree of compulsion should be used where feasible. There was a recognition of the requirement for both education of researchers on the need to provide key information as well as measures to ensure compliance such as withdrawal of funding in certain circumstances.

R003 Provide structures (common APIs, shared services or connectors) to support the exchange of research information, but not a central reporting system

a. Do not create a single national reporting system.

b. Provide common APIs to source, not transformed, research information.
c. Any technical solution for data exchange should be straightforward and have low integration costs.
d. Provide a single point of deposit for research outputs.

There was little enthusiasm for a single research reporting system across the sector. Considerable investment has been made by RCUK funders in systems such as ROS and Research Fish, which are now well-established. BIS is currently developing the Gateway to Research system. Also many larger institutions are making considerable and longer term investments in CRIS systems. Hence, any proposed solution should work within this framework. There was much stronger interest, both from institutions and funders on being able to harvest research information, in source rather than transformed format, and measures that could simplify this such as common API definitions. There was concern, particularly from institutions on the costs of integration with a national system, and a clear need for benefits for them. The solution should also be suitable for institutions with both large and small research budgets.

R004 Increase the quality and timeliness of research information across the sector

a. Improve quality control of research information.
b. Reduce human effort and increase automation in collection and processing of research information.
c. Implement administrator workflows to reduce possibility of human error.
d. Enable researchers to view and correct their own research information.
e. Use of shared services for validation and quality control.
f. Enable institutions to collect and validate research information prior to submission to funders.
g. Enable on-going reporting of research outputs to support ad hoc reporting by funders

Data quality was seen as a major issue by most stakeholders. In particular, there was a desire for further automation of processes, and automated or semi-automated validation. Bulk upload from institutions was in most cases seen as preferable to manual entry of information by researchers or research office staff into funder systems. In particular, cross-system synchronisation can be used to validate data. Enabling institutions to validate information prior to submission to funders will also result in an increase in data quality. An important caveat to this is that some funders value more qualitative information that researchers provide, particularly around the wider impact of their research. Hence they are keen to maintain a relationship with researchers working on their grants.

Where manual entry is required, this should be supported by validation workflows. Researchers and administrators should be able to log in, review and correct their entries as appropriate. Information that is collected for a specific purpose is often of higher quality, so information collection should be reduced to essential pieces of information. There is a requirement both in institutions and funders for ad hoc reporting on a short term basis. Institutional research offices need to respond to reporting requests. Funders need to respond to information requests from BIS and other government bodies.

R005 Facilitate the flow of information between internal institutional systems and external systems (e.g. funder systems) in CERIF format

a. Integrate internal systems with CRIS to reduce re-keying and enable institutions to collate information for reporting.
b. Enable bulk upload of data from CRIS systems to funder systems.

Research information is currently spread across multiple systems. There is a strong requirement from institutions to improve the interoperability and synchronisation of internal systems related to research such as finance, human resources, institutional repositories and CRIS systems. In order to compile reports for funders, there is often a large amount of re-keying of data, resulting in omissions and errors. This results in an inefficient and expensive process. There is wide acceptance that institutions should, where feasible, collect and upload research information from researchers to funders.

R006 Enable institutions to more effectively consume and re-use research information (e.g. for benchmarking and management information, portfolio management, collaboration, compliance monitoring, communications)

a. Support data harvesting of data from multiple funder systems.
b. Support for data harvesting from other institutions.
c. Provide benchmarking tools.
d. Provide ability to analyse data in different ways (e.g. according to department, collaborative network).
e. Provide support for communications and marketing.

There was strong demand from institutions for business intelligence tools to provide management information. Institutions with CRIS systems can already generate internal reports. However, institutions were keen to harvest data from external sources such as funders to be able to benchmark their performance against other institutions. There was a requirement to carry out this analysis in different ways such as by departmental or by research area. There was also strong interest in tools to support research portfolio analysis, strategic planning, analysis and promotion of collaboration, and compliance monitoring. There was demand to more easily generate full CVs for researchers, both for use for internal management purposes such as staff development, as well as to simplify the process of submitting grant proposals. There was an interest in automating the process of uploading and publishing research information to websites to support external communications and marketing.

R007 Support benchmarking and portfolio analysis across research funders

a. Enable funders to harvest research information from other funder systems.
b. Support benchmarking across funders.
c. Support research portfolio analysis across funders.
d. Support measure of long term impact of research.

Funders also had a strong interest in benchmarking their performance across other funding organisations. In order to support this analysis, agreed definitions of research impact metrics are required. Charities also had an interest in measuring research quality but were more strongly motivated by qualitative information that could be used for raising awareness to support fundraising activities.

R008 Provide appropriate data governance, transparency and security when collecting, sharing and reusing sensitive research information

a. Ensure compliance with data protection legislation.
b. Protect the confidentiality of commercially sensitive data.
c. Maintain trust of researchers in the use of the data.
d. Provide retention policies to support long-term monitoring.
e. Provide rigorous validation of data before release into the public domain.

Research information contains sensitive information relating to individuals as well as commercial organisations. Thus processes are required to protect confidential information. This can be supported by appropriate access management and security within systems. Use of data should be compliant with data protection legislation. Researchers were particularly concerned about how the data they provide might be used. Thus clear terms of use should be provided by institutions and funders requesting this information. Reporting on research information was generally based on aggregated or anonymised data, rather than on specific individuals. Rigorous workflows are required to support the publication of data, to ensure both that confidential data is not released, and also that the data is of high quality.

5. Phase 2 goals and options

This section details the end goals beyond the UKRISS project that the requirements and drivers indicate are desired by the sector. We will define three areas of general work where there is demand, and their relationships to the requirements.

We will then go on to present three distinct options for the work that could be carried out during Phase 2 of the UKRISS project. Each of those options is focussed on one particular aspect of the work required in the sector. The options are constructed from component parts which themselves arise directly from analysis of the requirements and drivers, and these are documented in detail in **Section 6** and in **Appendix J**. Each of the options also contains an analysis of the benefits for stakeholders and a supporting business case for consideration.

In **Section 7** we make some concluding remarks and a specific recommendation as to which of the project options we believe should be taken forward. This includes recommendations from both the project team and the steering board.

5.1 Business goals

Based on an analysis of the drivers and requirements, as well as analysis of relevant technologies, the project team has identified three key business goals.

G001 (Modelling) Specification, standardisation and adoption of a core CERIF profile for reporting of research information in UK HEIs (R001, R002)

G002 (Reporting infrastructure) Implementation of a national CERIF connector and associated shared services to facilitate the exchange of research information between IT systems within institutions, funders and statutory bodies (R003, R004, R005, R008)

G003 (Benchmarking) Provision of benchmarking tools that enable comparison and analysis of research information generated by multiple organisations for management information purposes (R006, R007)

Goal G001 defines a core information profile and CERIF serialisation for exchange of research information, addressing requirements R001 and R002. Goal G002 involves providing a robust exchange mechanism for moving the information between organisations, addressing requirements R003, R004, R005 and R008. Goal G003 provides a tool for exploiting the exchange of research information to generate management information and addresses requirements R006 and R007. Thus the three goals provide a framework for promoting and exploiting the exchange of research information across the sector.

This section of the report will provide a set of options for investigating the feasibility of achieving one or more of these goals. The proposed work in Phase 2 will include analysis of the most relevant technical, non-technical and business issues.

5.2 Project options

This section defines the options for UKRISS Phase 2 and their alignment with the overall business goals. Only one of the options can be carried forward into the second phase of the project. Nonetheless, the options presented here are idealised cases based around themes of work. The precise details of the work to be carried out will be presented in a separate project plan document on the commencement of Phase 2. The plans will be subject to review according to available resourcing as well as to input from JISC and the Steering Board during the work itself.

Three options for Phase 2 are presented as follows:

1. **Focus on modelling:** CERIF modelling to address only business goal G001. The focus of this work is on describing the core information profile and achieving a full serialisation of this profile in CERIF. The core work could be extended to include work on integrating CERIF models for HR data within institutions.
2. **Focus on benchmarking:** This option addresses business goals G001 (to a lesser degree) and G003, with a focus on benchmarking as a main outcome, with sufficient modelling work to develop a core information profile and its serialisation in order to support that benchmarking.
3. **Focus on reporting infrastructure:** This option addresses business goals G001 (to a lesser degree) and G002, with a focus on a proof-of-concept research reporting infrastructure and process, including the feasibility of providing shared services. There will be sufficient modelling work to develop a core information profile and its serialisation in order to support the reporting requirements.

In order to implement these three options a set of components has been defined, which are summarised in **Table 5.1**. Some components can have Core and Advanced levels, in order to be adaptable to time and resource constraints, and are referenced by unique identifiers. A detailed description of each of the components is contained in **Section 6.**

Seq	Component	Level	Code	Description
1	Information Profile	Core	1C	Describe the **Core Information Profile** aligned to the reporting requirements of a sample set of funders
		Advanced	1A	Expand the core information profile and perform a gap analysis on information available from institutions
2	Serialisation	Core	2C	Map the core information profile to CERIF
		Advanced	2A	Make recommendations regarding CERIF extensions, and port the CERIF serialisation to the VIVO data model.
3	Reporting infrastructure	Core	3C	Deploy and evaluate CERIF connector for point-to-point information transfers between institutions and funders.

		Advanced	3A	Extend the infrastructure of 3C to incorporate transfers via the JISC Nexus cloud ESB[25].
4	Benchmarking	Core	4C	Develop user interface enhancements and indexing enhancements to the SolrEyes software to support benchmarking
		Advanced	4A	Extend the dataset and processes of 4C
5	Standard Terms	Core	5C	Define standard terms and common elements
6	Institutional model	Core	6C	Definition of an institutional data information set and CERIF model
		Advanced	6A	Make recommendations regarding CERIF extensions, and explore further available institutional data

Table 5.1: Components for phase 2 (see Section 6)

These components have inter-dependencies which imply causal relationships between them, and these are as shown in **Figure 5.1.** Each option will highlight the part of this set of components that it will address.

[25] An Enterprise Service Bus (ESB) refers to a piece of software that enables communication between multiple business applications. All direct contact between the applications takes place via the ESB, using a standard set of messages. An ESB hosts a large collection of services. Services deal with the routing of the messages, as well as providing commonly used functions such as data format transformation, and queuing and buffering of data to support different processing speeds.

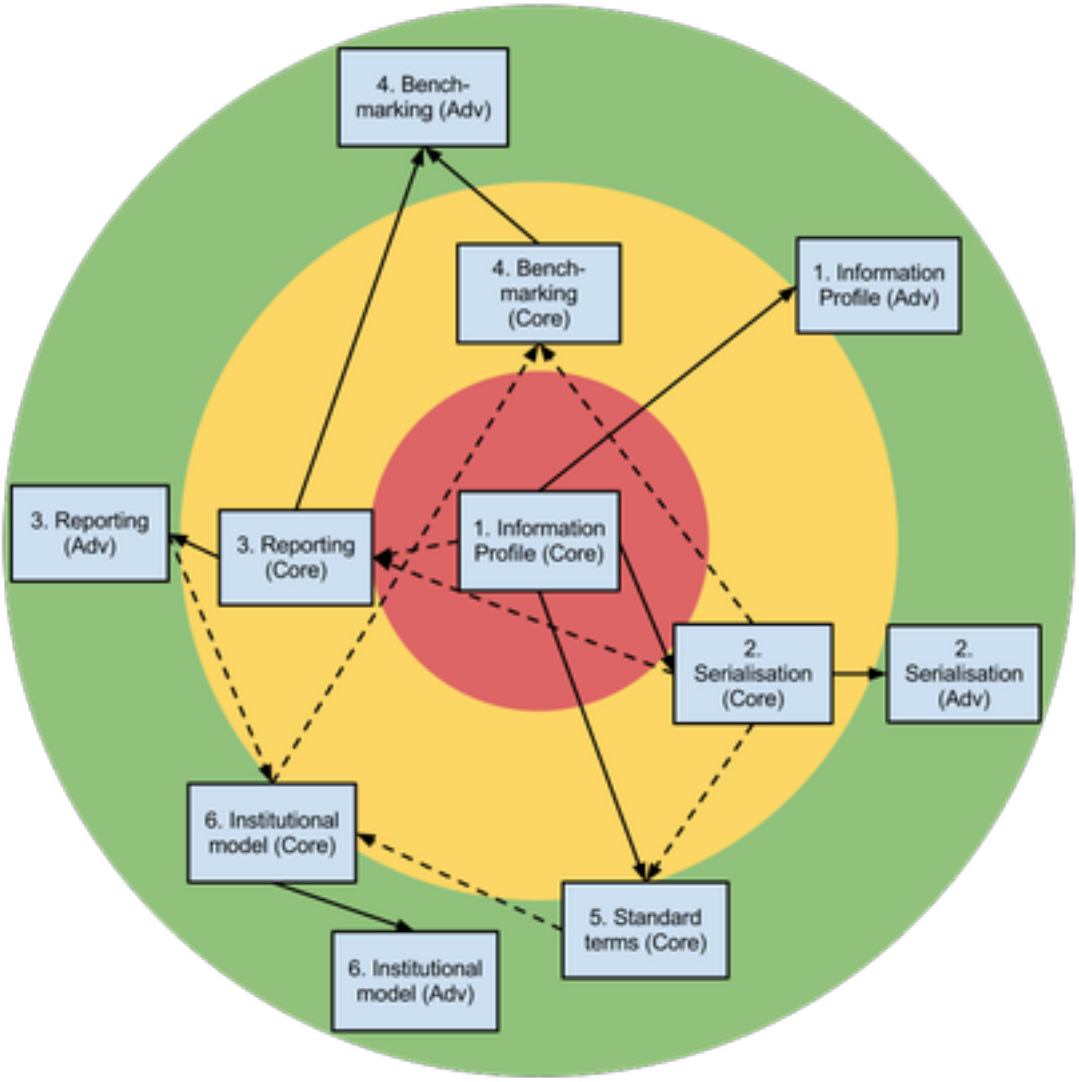

Figure 5.1: inter-dependencies of the 6 components from which the Phase 2 options will be derived. The solid arrows define 'hard' dependencies (one component is fully predicated on another), while dotted arrows define 'soft' dependencies (one component can benefit from the outputs of another).

In **Sections 5.3, 5.4** and **5.5,** we describe the options in more detail, referencing the components defined in **Table 5.1,** together with a supporting business case.

5.3 Option 1: Focus on modelling

5.3.1 Description

The main objective of the modelling activity is to describe a core information profile for research reporting across the sector and a set of mappings to enable unambiguous generation of the core information profile in CERIF. The core information profile would contain a minimum set of fields to be collected by all

funders and statutory bodies across the sector, to enable sector-wide comparison. It can be defined by examining the information that funders currently collect, and looking for commonalities across multiple such organisations; this will be a passive process - there is no attempt within this option to effect any change in the funder reporting processes. See **Section 6.1** for full details of this component part of the work.

The core information profile can be extended in two ways. Modelling of institutional data including information about organisational structure and HR data can be included to provide a consistent way for institutions to collect and collate reporting information for submission to funders, and to perform analysis of the research information according to organisational unit or activity. Secondly, more detailed work can be performed to define dictionaries and common terms required for the core information profile.

The modelling option has three main aspects. The first is to collect sufficient information regarding reporting requirements from a wide range of funders and statutory bodies that a set of core information fields can be determined. For the institutional data, additional information will need to gathered from a sample set of HEIs. The second is to perform the CERIF mappings and to support standardisation of the profile within euroCRIS. The third is to drive adoption of the core profile through consultation and promotion within the sector.

5.3.2 Phase 2 work plan

This option focuses on the information and CERIF modelling required by many of the components proposed for UKRISS Phase 2, but does not develop any software or services around the models.

Figure 5.2: Essential modelling subset of the components

The tasks carried out by this option will be:

- Define the core information profile (1C). This includes collection of reporting requirements from a wide range of funders and statutory bodies
- Serialise the core information profile to CERIF (2C)
- Define an institutional data information set and CERIF mapping (6C)
- Perform a gap analysis between institutional data modelling requirements and CERIF (6C)
- Produce a set of recommendations for standard interpretations of a subset of all useful terms (based on the work achievable in the time available) (5C)
- Proposals for standardisation to be submitted to euroCRIS.

By the end of the UKRISS project we will therefore have developed the core information profile and and the institutional data information set and their CERIF serialisations. We will also have carried out a

consensus-building exercise on standard interpretation of a sub-set of useful terms. The final output of the project will be a set of documented proposals for standardisation/harmonisation options; there will be no functioning outputs such as software.

Month	1	2	3	4	5	6	7
1C.WP1							
1C.WP2							
1C.WP3							
2C.WP1							
5C.WP1							
5C.WP2							
6C.WP1							
6C.WP2							
6C.WP3							
Evaluation							
Sustainability							

Table 5.2. Timescales for implementing option 1

5.3.3 Dependencies

The success of the work is dependent on:

- Obtaining information on reporting requirements from a wide range of funders (including charities), institutions and statutory bodies
- Support from euroCRIS to adopt the proposals made by the project into the standardisation activities
- Collaboration with related UK initiatives on CERIF, including Gateway to Research, JISC and RCUK funded projects, to achieve agreement on the CERIF mappings of the core information fields
- Willingness of organisations to collect information fields that are not mandated by their current reporting requirements
- Readiness to adopt the core profile, particularly by CRIS vendors and funder systems.

5.3.4 Strengths and weaknesses

Definition of a core information profile and CERIF mappings would enable consistent reporting of research information across a range of common fields. This would reduce the amount of duplication required to generate reports for different funding organisations, particularly for co-funded projects. The profile would enable organisations to compare and benchmark their performance against other organisations in the sector.

The main weaknesses of addressing only the information aspects is that data quality and data governance issues are not addressed, particularly relating to requirements R004 and R008. Thus exchange of research information would need to be carried out on an ad hoc basis, possibly using different interfaces. There is no common mechanism for ensuring the integrity of the data, so each organisation would be responsible for performing its own validation and quality control.

The core information profile itself is an important enabler. It does not bring any tangible benefits without additional tools to demonstrate value to stakeholders. Thus it may be difficult to gain acceptance and adoption of the profile without additional applications that can exploit it.

5.3.5 Benefits and beneficiaries

The principal beneficiaries of a core information profile would be researchers, institutions, funders, statutory bodies, CRIS vendors, the euroCRIS community and JISC. This option addresses a primary requirement R001, which is a wide engagement with stakeholders across the sector, including those that do not currently engage directly with CERIF. Such an engagement would already lead to greater awareness and uptake, promoting increased information exchange.

Benefits for institutions

The amount of duplicate reporting and reformatting will be decreased since a common set of information can be repurposed for multiple reporting requirements. Institutions can harvest data from multiple sources, including funders, statutory bodies and other institutions in the knowledge that it will contain a standard set of information fields, without the need for additional processing. Given access to data from other institutions and suitable software tools, there is the potential to perform benchmarking, portfolio analysis and develop collaboration tools through analysis of a comparable set of information.

Benefits for researchers

For researchers, reporting processes will be simplified. Regardless of whether researchers are reporting internally or directly to funders, a core set of the information fields requested will be the same, resulting in a reduction of time and effort.

Benefits for funders

Given a core information profile, much of the information collected by funders will be comparable, enabling funders to exchange information about their portfolios to reduce duplication, to perform benchmarking, and to generate standard reports (e.g. for government) based on the core information profile.

Benefits for vendors

Institutional CRIS vendors have a strong interest in standardisation of data formats to enhance the functionality of their existing products, whilst at the same time reducing the need for adapting to changes in core standards. Providing a persistent core information profile would give vendors much greater certainty when developing their products, whilst also providing the potential to develop additional reporting and benchmarking tools to exploit shared information.

Benefits for statutory bodies

Organisations such as HEFCE and HESA collect a wide range of information from institutions, which places a considerable administrative burden on institutions as well as the statutory bodies themselves. Aligning the statutory reporting requirements with a core information profile would greatly simplify collection of information, since it would reduce the need for bespoke information collection. Repurposing of the information is also likely to result in higher data quality, since each submission of the information would require validation.

Benefits for the euroCRIS community

There is widespread international interest in the development of CERIF in the UK, which is widely regarded as internationally leading. Definition and standardisation of such a profile in the UK would provide a blueprint for other countries. Given the increasingly international nature of research activities, international adoption would also have considerable future benefits for the UK in terms of cross-border

information exchange.

Benefits for JISC

Work on the CERIF mappings for the core information profile would draw on the work carried out by other JISC projects in the Research Information Management programme that have implemented specific use cases for information exchange. Unifying this work in a core profile would serve to enhance the overall impact and adoption.

5.3.6 Costs and benefits

The initial work on the core information profile can be carried out within the UKRISS project, and the cost is therefore the remainder of the project budget. Additional costs incurred beyond the end of the project would be to carry out standardisation of the profile within UKRISS, and some indication of the work required is provided in **Section 5.3.9**.

The RMAS project carried out standardisation of 163 fields in the CERIF profile, which involved attendance of five euroCRIS meetings at a total cost of £10k, and this may be indicative of the cost of extra work beyond the end of the UKRISS project.

The benefits to the sector depend on the rate of adoption of the core profile beyond the end of the project. A strong business case for CERIF adoption was made by the JISC funded report in 2010[26]. The work carried out by this project is also an enabler for many other developments within the sector, so the true benefits may not be realised on a short timescale.

5.3.7 Scalability

The scope of the core information profile would need to be restricted to a set of fields on which wide consensus could be obtained. The profile could be extended at a later stage when the principle and value were established. Similarly there are areas in CERIF such as definitions of non-publication outputs which in some cases still require further clarification. It might be necessary to omit these until further work could be carried out to obtain agreement.

5.3.8 Evaluation

The evaluation for this option would primarily be based on feedback from stakeholders on the potential value and likely adoption of the core information profile and extensions. A wide range of stakeholders will be engaged early on in the process, and will be consulted during the development of the profile. Elements of the profile will be published on a website in order that remote feedback can also be collected. The UKRISS Steering Board and the extended UKRISS Phase 1 study participants, including CRIS vendors, provide a wide cross-section of the sector to provide feedback. Through euroCRIS, we will obtain critical feedback from both UK and international experts on the viability of the profile and to assist in identification of gaps or potential technical deficiencies.

5.3.9 Sustainability

The planned work would follow the overall steps:

[26] JISC (2010) Business case for the adoption of a UK standard for research information interchange: http://www.jisc.ac.uk/publications/reports/2010/businesscasefinalreport.aspx

1. Definition of the core information profile and CERIF serialisation
2. Establishment of a governance group to oversee the on-going development of the profile. This group would include key stakeholders such as the funding councils and statutory bodies. This might be chaired by a sector-wide body such as UCISA, ARMA or carried out within euroCRIS
3. Standardisation of the core information profile. This would ideally be carried out within the timescales of UKRISS, or could be taken over by euroCRIS or other appropriate body at the end of the project
4. Promotion and marketing of the profile. This might involve further funded work from JISC to embed the profile within the sector in collaboration with the governance group
5. Engagement of institutions, funders and CRIS vendors, in order to encourage discussion around the profile with the view to getting a wide technical and policy acceptance (and bearing in mind that the profile will no doubt change during discussions)
6. Ultimate adoption of the core information profile and a process for its governance to support change over time.

The overall timescales would be as follows: Steps 1-3 should be completed by the end of UKRISS. Activity 4 will be initiated during the project would need to continue for at least year beyond the project, supported by JISC and any governance group. Activity 5 would come from initial engagement with the community in Activity 4, and would itself be a fairly substantial process requiring funding, either from JISC or from the Research Information community. Activity 6 would naturally follow from a successful Activity 5, and the long term governance of the profile would also have to be determined, but that is outside the remit of the UKRISS project.

5.4 Option 2: Focus on benchmarking

5.4.1 Description

The main aim of this option is to focus on the development of benchmarking tools to analyse research information harvested from multiple institutions. Preliminary work is required to define the core information profile and the required CERIF mappings from option 1.

This option will develop a set of open source benchmarking tools that exploit the core information profile to meet the requirements R006 and R007.

In parallel to this work, we will investigate the information sharing issues around ad hoc data transfers to support benchmarking, addressing requirements R004 and R008. This would include examining the quality of the data, the implications for the benchmarking metrics that can be produced (such as error ranges) and data governance issues, especially around the willingness of institutions to share their data and on what granularity (this has very close ties to the Snowball project). We will therefore engage stakeholders early in the process, to ensure that the tools and processes developed are appropriate and viable.

5.4.2 Phase 2 work plan

Figure 5.3: Essential benchmarking subset of all components

The tasks carried out by this option will be:

- Define the the core information profile (1C)
- Serialise the core information profile to CERIF (2C)
- Define generic benchmarking reports required at institutions and define a flexible mapping from CERIF to an indexable form to support those reports (4A)
- Develop user interface enhancements to the SolrEyes software to support benchmarking and an extended dataset (4A, 4C)
- Towards the end of the project, incorporate the core information profile as per the serialisation from 5C (2C).

Note that we omit the continuous integration of data from institutional systems, as this relies on the reporting component (3C).

By the end of the UKRISS project we will therefore have developed a first version of the core information profile and its CERIF serialisation, and using this data model have built some proof-of-concept, but relatively advanced benchmarking tools which take into account the policy issues around data sharing.

Month	1	2	3	4	5	6	7
1C.WP1							
1C.WP2							
1C.WP3							
2C.WP1							
4C.WP1							
4C.WP2							
4C.WP3							
4A.WP5							
Evaluation							
Sustainability							

Table 5.3: Timescales for implementing option 2

5.4.3 Dependencies

Key dependencies of the work on benchmarking are as follows:

- Definition and adoption of a core CERIF profile for reporting research information to enable a wide range of information and indicators to be compared across the sector
- Definition of benchmarks and metrics to be used in the tool. The work in the Snowball project[27] would be relevant. The tool could be flexible to allow the user to define additional metrics and comparators
- Development of the benchmarking tools is dependent on obtaining research information from multiple funders and/or institutions of sufficient quality to enable realistic results to be obtained
- Engagement from end users at institutions to assist with the specification and evaluation of the tools.

5.4.4 Strengths and weaknesses

Development of proof-of-concept benchmarking tools based on information in the core information profile will demonstrate both the value of institutions and funders of the profile as well as promoting the value of research information exchange across the sector. In particular, CRIS vendors could adapt and extend the tools developed in the project to provide additional value for institutions.

Providing only benchmarking tools does not address the requirement to provide a single point of contact for exchanging research information. Thus all exchanges would need to be point-to-point using ad hoc interfaces. For information with larger bodies such as ROS, Research Fish or statutory bodies, this does not present a huge burden. However, it does present a major barrier to cross-institutional information sharing or submissions to the very large number of smaller funders. There is no option for including quality control, standardised reporting (i.e. responses from the recipient) or compliance monitoring.

Harvesting information for benchmarking would also be complex. This could be done by harvesting public research information across multiple funder systems, by bilateral information exchanges or from institutional repositories. National systems such as Gateway to Research may also provide a subset

[27] Snowball Project: http://www.snowballmetrics.com

of the information required. Timeliness of the research information may also be an issue. Published research information may relate to different time periods making it difficult to compare.

5.4.5 Benefits and beneficiaries

In this section, we consider the incremental benefits of developing the benchmarking tools in addition to those benefits of a core information profile described in option 1.

The main beneficiaries of the benchmarking work would be institutions, funders and CRIS vendors.

Benefits for institutions

The benchmarking tools would simplify the task of generating management reports for Research Office staff. Ideally the benchmarking tools would be integrated with institutional CRIS systems to provide full access to internal data. The primary beneficiaries would be institutional research managers. The benchmarking tools would enable far more effective use of resources, monitoring of research activities, portfolio analysis and collaboration.

Benefits for CRIS vendors

Vendors would have considerable interest in proof-of-concept benchmarking tools, which could be enhanced and integrated with existing products to provide additional value for institutions.

Benefits for funders

Demonstration of benchmarking over a core information profile would provide funders with a mechanism to carry out their own funder-funder benchmarking and portfolio analysis.

5.4.6 Costs and benefits

The costs of the proof-of-concept benchmarking tool and core information profile are covered by existing UKRISS funding. Further development of the benchmarking tools could either be performed by the HEI community, through additional grant funding or taken up by vendors for integration into their CRIS product offerings.

Phase 2 of the project will aim to identify and quantify potential benefits of use of benchmarking tools on institutional efficiency. The overall objective of benchmarking is to drive research quality and process improvement, and any gains are likely to be difficult to quantify in purely financial terms.

There will also be general costs in organisational change in adopting any new software and practices, although these should be one-off costs.

5.4.7 Scalability

Benchmarking tools would need to be able to import and process large quantities of data harvested from funders or other institutions. This may place a large burden on existing CRIS infrastructure. In some cases, it may also be more cost effective to hold research information in external cloud storage in order that information can be processed more efficiently and to reduce the required investment in internal servers. Phase 2 of the project will investigate the overall scale of the data, the processing overheads and the resulting infrastructure costs to support effective benchmarking.

5.4.8 Evaluation

The success of this option will be determined by the effectiveness of the benchmarking tools and their utility for research managers and research office staff. Technical evaluation will consider the range of potential metrics can be provided and their effectiveness in meeting the institutional strategic requirements. We will evaluate the impact of data quality of the robustness and reliability of the tools and aim to provide confidence values around the metrics that are produced.

Usability of the benchmarking tools will be tested through a formal user evaluation that evaluates both the ease of configuration of the tool for research office staff as well as the quality and presentation of the reporting information.

5.4.9 Sustainability

There will be two main outputs from this option:

1. An Open Source set of benchmarking tools
2. Initial work on the core information profile and its CERIF serialisation

The software will be made available on a public source code repository (e.g. github), but this is insufficient to ensure take-up. In addition we will engage the research information community (e.g. through ARMA) to encourage usage and further development of the software, with a view to building a community around it in the long-term (not within the time-frame of the project). There may also be routes to sustainability through vendor or JISC-provided services built on the processes and software, which can be sustained through a traditional business model; through engagement with the stakeholders we will investigate the interest in the community for such services.

Meanwhile the core information profile and its serialisation would need to be further developed, most likely with grant funding initially, with a long-term view to constructing a governance organisation (e.g. euroCRIS) who can take it forward in the future.

5.5 Option 3: Focus on reporting infrastructure

5.5.1 Description

This option focuses on developing a proof-of-concept research reporting infrastructure and the processes and information required to go along with it. The project will make use of the RMAS infrastructure which provides connectors for various institutional systems - such as HR - and allows the data to be aggregated and converted into desired formats. We will therefore include the work required to develop the core information profile and a representation of that profile in CERIF, which will then be used as the common format for the research information aggregated therein.

The aim of the infrastructure is to provide a robust mechanism for the exchange of research information in CERIF format, where institutions and funders can connect through a central point of contact (an Enterprise Service Bus (ESB) in the cloud) to retrieve the desired reporting information. This infrastructure will be deployed in a cloud environment so that it is possible for multiple institutions and multiple funders to all connect to each other through the service, providing data consistent with the core information profile, simplifying the reporting requirements, and reducing effort within institutions.

The project will approach this challenge by considering a test project, which will be constructed to include a variety of important entities and information, such as researchers, publications, research data, etc, and

examine how the project's reporting requirements evolve over its lifetime. The project will also be cross-institutional and co-funded, to ensure that the complexities of such reporting are properly understood. Full details of the approach are available in **Section 6.3**.

As well as considering the technical aspects of the connection, this work will investigate the data governance and data quality issues around sharing data via the connector. We will consider to what extent the quality, compliance and data privacy issues can be addressed by additional shared services. For example, a data quality service could cross-check data with publicly available data sources to provide additional validation that could not cost-effectively performed by individual institutions. A further option would be to provide a national data warehouse that captured all exchanges of research information.

Business models for the reporting infrastructure and for offering shared services will be proposed and evaluated by conducting interviews with institutional and funder staff responsible for management of research information reporting activities.

5.5.2 Phase 2 work plan

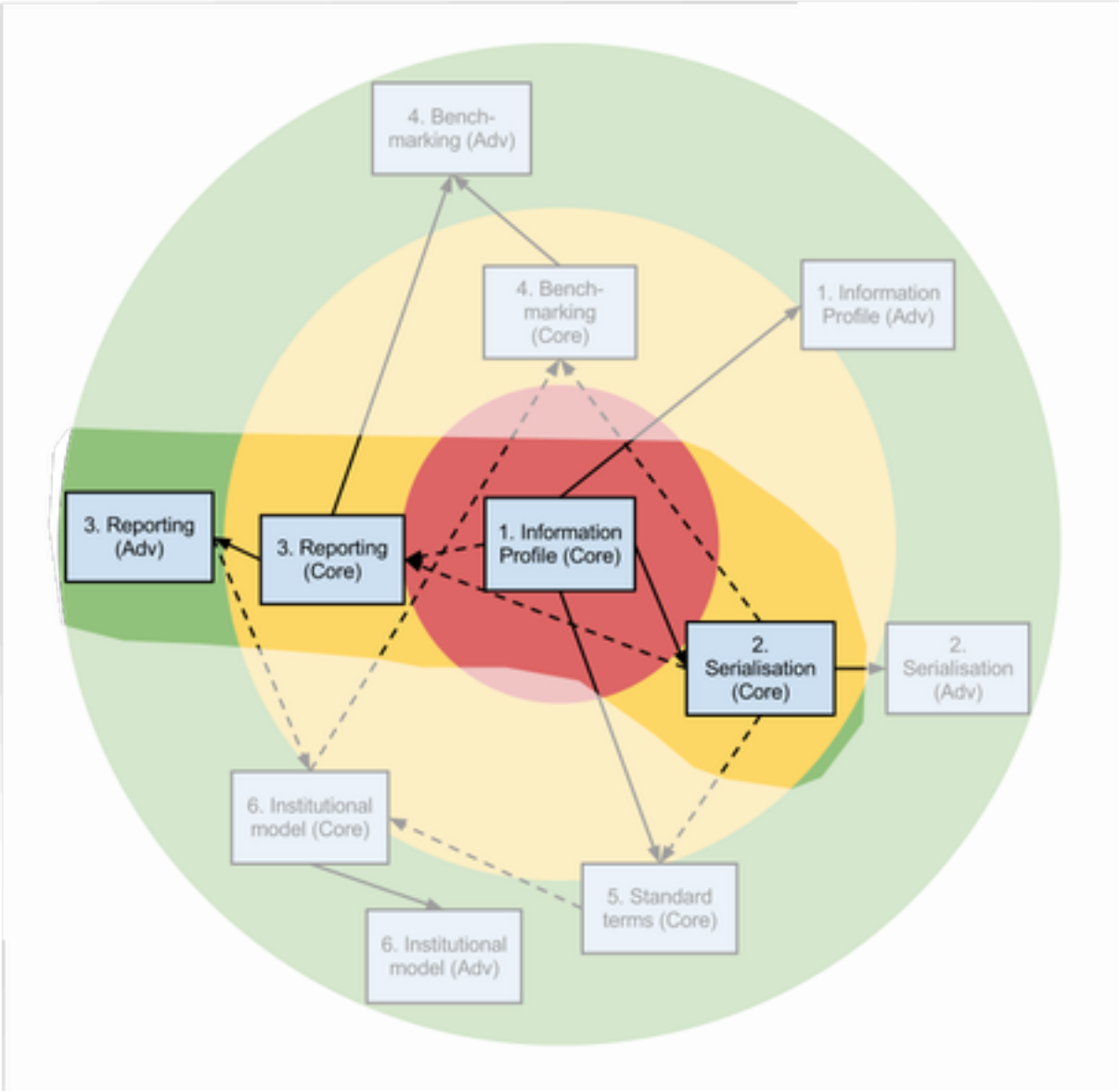

Figure 5.4: Essential modelling, national connector and benchmarking subset of components

The tasks carried out by this option will be:

- Define the core information profile (1C)
- Serialise the core information profile to CERIF (2C)
- Deploy the RMAS infrastructure at one or more UKRISS partner institutions (3C)
- Deploy and evaluate CERIF connector for point-to-point information transfers between institutions and funders (3C)
- Deploy the reporting infrastructure on the JISC Nexus cloud (3A)

By the end of the UKRISS project we will therefore have defined the first version of the core information profile and its CERIF serialisation. We will also have deployed the RMAS infrastructure both at one or more institutions and in a cloud environment, and modelled the processes around project reporting requirements and information. The outputs of the project will be documentation of those processes, initial RMAS configuration of data connectors, and a feasibility study of rolling the infrastructure out at a national

level, including a proof-of-concept technical set-up.

Month	1	2	3	4	5	6	7
1C.WP1							
1C.WP2							
1C.WP3							
2C.WP1							
3C.WP1							
3C.WP2							
3A.WP3							
3A.WP4							
Evaluation							
Sustainability							

Table 5.4: Timescales for implementing option 3

5.5.3 Dependencies

This approach depends on the following key elements:

- Development of core information profile and a CERIF representation for reporting research information. Thus a single report could be routed to multiple funders via the ESB.
- Availability of a cloud ESB connector and resources to be able to configure it, both for use in a proof-of-concept as well as in a production service. It is proposed to use the JISC Nexus ESB for project work. A commercial decision would then be made by JISC on the future hosting of the service.
- Development of additional shared services via the ESB (e.g. quality control, reporting, data warehousing, bibliographic services) to generate interest from institutions and funders, and to make such a service commercially viable. The potential for such services can be identified in the project and presented as part of the final business case at the end of the project.

5.5.4 Strengths and weaknesses

The national connector option offers the potential to simplify exchange of information between organisations and provide additional shared services. This approach has the potential to contribute to technical solutions for issues such as data quality and data privacy by providing additional shared services.

The main weakness of the approach is that it requires deployment and adoption of a national infrastructure, which will result in costs for organisations that participate.

5.5.5 Benefits and beneficiaries

The main beneficiaries of a national connector are institutions, funders, CRIS vendors and statutory bodies.

Benefits for institutions

The national connector enables institutions to simplify their upload and harvesting of research information by providing a single interface for information transfer. Using a national connector in conjunction with an internal bus will reduce significantly the need for re-keying of data. Implementing an information filter at the end point of the CERIF connector in the institution enables information to be collated once from

internal systems and then filtered to provide reports to multiple recipient organisations. The filter would also provide a robust mechanism to prevent unwanted disclosure of confidential information such as personal or commercial data.

Use of the national connector greatly simplifies the infrastructure requirements on the institution in responding to information requests from multiple organisations.

The work done in this option will also enable future benchmarking and other analytical approaches which will improve management information and thus lead to better strategic decision making.

Benefits for funders

The national connector would simplify the connections of funders to UK HEIs. Large funders potentially need to receive data from over 150 UK institutions. Additionally funder-to-funder transfer of information would be greatly simplified.

Provide data validation services to ensure the quality of data being exchanged. The RMAS project found that over 30% of research information stored within partner institutions was either incomplete or was erroneous[28]. Thus data validation services that compare data across multiple sources for inconsistency and incompleteness would be of great value.

Statutory bodies

The benefits for statutory bodies are similar to funder benefits. Statutory bodies in particular have strict compliance and quality controls, which results in rejection of incomplete or incorrect information. Such communications could be carried out far more efficiently via the ESB and could be assisted by additional shared services.

CRIS vendors

Institutional CRIS systems would provide the main endpoint for the national connector. Thus there is a potential for CRIS vendors to provide services to support research management, such as cross-organisational benchmarking based on timely and high quality data, as well as providing a gateway to external information resources for researchers.

5.5.6 Cost-benefit analysis

In this section, we present a simple cost-benefit analysis to compare the deployment and running costs of a national connector to the project cost savings based on time savings resulting in reduced duplication of reporting.

The deployment of the national connector at an institution would involve:

- Setting up an institutional server
- Setting up an internal RMAS connector
- Configuring the connection to the national connector
- General cost of institutional change of software and practices.

This would be a relatively minor deployment cost per institution, although the cost of institutional change is difficult to estimate. Nonetheless, all of these costs are one-off investments.

[28] RMAS Benefits Analysis 2012: http://www.rmas.ac.uk/pdfs/RMAS-Benefits-Summary.pdf

The recurrent costs would include:

- Server maintenance/software updates
- On-going data validation
- Running the infrastructure locally or subscribing to a cloud-hosted service.
- Data transfer charges for moving data in and out of the cloud.

The estimated annual costs would be in the order of a few tens of thousands, on average, and would be split between IT infrastructure (and support from that department) and data validation/administrative personnel. Use of a cloud-based service could drive this figure down, by utilising the economies of scale brought by such an approach.

The efficiency savings would include:

- A small amount of time saved per academic per project
- A significant amount of time saved per administrator

This efficiency improvement could be of the order 1 FTE per organisation. Ignoring initial set-up costs, the overall cost saving to the sector is the difference between 1 FTE of an administrator and the overall annual running costs of the service, per HEI.

Projected rate of adoption

Year 1: 2013/14 – 0 new adopters

Pilot sites at Kings, Brunel and Exeter would refine the proof-of-concept developed during course of UKRISS project. Development of a production solution would be carried out by engagement of a commercial partner and/or JISC Nexus.

Year 2: 2014/15 – 3 early adopters

Following dissemination of results from pilot sites, we project two other HEIs and one funder would employ the UKRISS solution, in addition to the three pilot sites. If the JISC service model in the Cloud is pursued, we expect this to become operational in year 2.

Year 3: 2015/16 – 5 new adopters

Following dissemination of results from pilot sites, we project that three new HEI's and two funders will employ the UKRISS solution.

For this trajectory, the project delivers recurrent productivity savings of many hundreds of thousands of pounds per annum by year 3.

5.5.7 Scalability

The main element infrastructure dependency for this option is the cloud ESB. The ESB can be scaled according to the required number of connecting organisations and data transfer volume. Thus it would be possible to progressively add organisations to the system, with a proportionate deployment of additional servers and storage to support the service. The staffing costs to maintain the system would be largely independent of the number of subscribers.

5.5.8 Evaluation

The project would provide a technical evaluation of sending reporting messages over the cloud ESB connector to uncover any potential technical barriers. We will also evaluate the writing of business rules for routing the research information through the ESB. We will investigate capacity and scalability issues that might arise when adding multiple institutions to the connector.

A major part of the evaluation will be non-technical. We will demonstrate the proof-of-concept to institutions, funders, statutory bodies and CRIS vendors to determine the potential business opportunity for such a service and its economic viability. This will include the scope for providing additional shared services, which might be provided on a fully commercial basis. For this reason, we will consult stakeholders from the outset of this project, in order to determine which modes of deployment and business models would be viable.

We will also evaluate the potential of this approach to address the data quality and data governance requirements R004 and R008, through both technical measures as well as through obtaining feedback from stakeholders.

5.5.9 Sustainability

The reporting infrastructure could potentially be extended to provide a national service for the UK HE sector. During the sustainability phase of the project, we will discuss the potential for such a service with JISC and candidate service operators. This activity will be supported by technical, non-functional and business intelligence gathered during the proof-of-concept work.

Sample timescales for development and adoption of the system would be:

- Implementation of a full national pilot system (1 year to mid-2014). This would involve deployment with a larger number of invited early adopter funders and institutions
- Development of a commercial service offering with potential for all UK HEIs to subscribe for the service (end 2014)
- Widespread deployment across the sector (end 2015).

The move to an adoptable service will likely require some funding to transition from a proof-of-concept technical infrastructure to a production-quality one, possibly via or in collaboration with a vendor/service provider. Roll-out across the sector, then, would depend on the perceived value of the service; this would be most significantly shown if the key research councils adopt this approach, so it will be critical to engage them in the transition process.

Meanwhile the core information profile and its serialisation would need to be further developed, most likely with grant funding initially, with a long-term view to constructing a governance organisation (e.g. euroCRIS) who can take it forward in the future.

6. Phase 2 components

This section goes into significant detail on the six areas of work which form “components” of the overall options for UKRISS Phase 2, as presented in the previous section. They are presented here as self-contained projects with their own approach/methodology, risks, timelines and workpackages, specified over the full 7 month period allocated to Phase 2. It should be noted that these are idealised cases, and in the second phase of this project they will be subject to some change based on their relationships to other components presented.

These components have been drawn from a more complete list of potential areas of work which have been derived from the requirements and drivers gathered during UKRISS Phase 1, and this full list can be seen in **Appendix J.**

	Component	Drivers	Requirements	Key Benefit
1	Information profile for reporting	D001, D002, D003, D004, D006	R002, R005	Enabler for further development around research information and to drive harmonisation of reporting requirements.
2	Serialising the core information profile	D002, D003, D004, D006	R001, R002, R004, R005, R006, R007	Provide a clear CERIF serialisation of the most important research information for use in related software systems and APIs, and to help drive harmonisation.
3	Enhanced intra-institutional, cross-institutional and institution-to-funder information exchange	D001, D002, D003, D004, D006	R003, R004, R005, R006, R007, R008	Streamlining the institution-to-funder reporting process, building on the harmonisation of the reportable information.
4	Shared reporting/benchmarking	D001, D003, D004, D005, D006	R004, R006, R007, R008	Improve management information and business intelligence through software and processes within an institution.
5	Standardising research information terms	D001, D002, D003, D004, D006	R001, R002, R006	Enabler for other work around research information, and the harmonisation of terms across the sector.
6	Institutional data CERIF models	D001, D002, D003, D004, D006	R001	Enabler for other work around research information, and in driving harmonisation across the sector.

Table 6.1: summary of components

The six components detailed in this section are broken down into **Core** approaches and **Advanced** approaches, and have some degree of inter-dependencies. **Figure 6.1** shows these inter-dependencies, with the Core and Advanced work separated out.

Figure 6.1: inter-dependencies of the 6 components from which the Phase 2 options will be derived. The solid arrows define 'hard' dependencies (one component is fully predicated on another), while dotted arrows define 'soft' dependencies (one component can benefit from the outputs of another).

6.1 Component 1: Information profile for reporting

Currently there is a lack of comprehensive information regarding reporting requirements to the various funding bodies, and a lack of consistent expression of reportable metadata. There is also a lack of consensus across all funding bodies regarding reportable metadata. Together these mean that reporting from institutions is complex, highly manual-work oriented, and involves a lot of duplication of effort, causing significant overhead to achieve compliance.

We will carry out an investigation into all of the requirements of the funders to find out what the sum total of all information requested is, and perform an information modelling exercise to define the maximum quantity of information that an institution will need to provide. This is a passive information gathering exercise, and does not presume to effect any change in the funder's reporting requirements.

There is also a lack of consensus/common agreement in the sector as to which is the most critical research information. This leads to difficulties in communication between similar systems and also difficulties in translating the data for re-use in other contexts (e.g. VIVO). The complexity of research information also presents a high barrier of entry for smaller organisations.

We will define the minimum amount of information required to transfer research information between institutions and funders, which is referred to throughout as the **core information profile**. This will include some CERIF modelling, the definition of the core set of metadata (a kind of Dublin Core[29] for research information), and any associated data dictionaries/taxonomies/semantics required.

The key benefit of this work is as an enabler for further development around research information and to drive harmonisation of reporting requirements.

6.1.1 Drivers for this component

D001 Improve business intelligence, management and due diligence through better information quality and reporting utility

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community through harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D004 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

D006 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.1.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- Encourage the funders (not just RCUK but certainly all UK funders, and potentially in the future EU and other funders) to be aligned on their reporting requirements and standardise the approach and data (for research reporting and statutory returns), to reduce duplication of effort
- Harmonise the information capture and descriptions of common entities (researchers, research outputs, etc)
- To capture and model as structured information a complete picture of research to allow an understanding of aspects such as collaborations, impact, outputs, students, professional development, which will be useful for management information

[29] Dublin Core: http://www.dublincore.org/

- Support submission to funders for returns through a common format (and in particular REF).

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.1.3 Approach and methodology

This section defines two pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra could be carried out in addition to the Core work.

Core Work

The primary task of this project is to gather current requirements from the funders regarding their reporting requirements, and de-duplicate and aggregate the fields that they are interested in as a whole.

In the core part of this project will will contact only RCUK funders and HESA regarding their requirements as they represent the main body of funded research in the UK and covers a wide range of subject areas.

We will take our findings from these bodies and document the total set of requirements as an information model. We will not attempt CERIF modelling the data at this stage, but we will look to the dictionaries/ profiles provided by CASRAI[30] to help us document the model.

Once we have understood the requirements of each of these bodies we will be able to look at which bodies require which fields, and determine where they overlap and where they differ. Based on the overlap we will be able to identify a core information profile which will effectively define the most essential research information in the sector, and which may be used by other UKRISS Phase 2 projects in modelling efforts.

We recognise at this stage that funder requirements will change over time, and that this component will only create a profile based on a “snapshot” of those requirements at the time of the study.

Advanced Work

To gain a wider perspective on the requirements of funders and related bodies we will include The Wellcome Trust, other charities, and other smaller funders into our requirements gathering exercise, as well as potentially the EU. We will also question each one regarding not only their current requirements, but also their “wish list” of other information they would like to be able to collect. This will be framed by also requesting the funders to justify why such information would be useful to them, to avoid unnecessary broadening of scope.

With this extra information, we will be able to carry out a basic gap analysis between what the ultimate reporting requirements might be and what institutions are capable of providing (based on information gathered from UKRISS project partners).

We will also aim to present our findings back to funders and other stakeholders with a view to beginning a discussion to drive agreement to harmonise across the sector. There is a strong need for political buy-in

[30] CASRAI: http://casrai.org/

at this stage, so the project will aim to bring the funders together around this artefact, and make the case in detail for harmonisation, and guide the group in discussions. This will also help us understand how we can manage changing funder requirements over time, and develop a framework for those changes to be incorporated into the information profile in future.

6.1.4 Feasibility and risks

Our main risks in this project are around the advanced work, and we are confident that the core work will proceed with minimal complications. Our primary mitigation tool is to use the definition of a **core information profile** to drive the focus towards the parts of the research information which are clearly of the most value to the community.

Risk: Conflicting with competing activities such as Gateway to Research

Likelihood (L): 3; **Impact (I)**: 1; **Factor (L X I)**: 3

The project is well aware of other work going on in this sector, and has previously carried out an extensive landscape study to allow us to discover efforts to which we need to pay attention during this work. We will be sure to look at the outputs of such projects, and ensure whatever we produce fits well with what they are doing, and plan to base our work on other projects, rather than attempting to cut across them.

Risk: Unable to drive funder agreement around a common information set

Likelihood (L): 5; **Impact (I)**: 2; **Factor (L X I)**: 10

While this is quite likely, it will impact the project only minimally as driving discussion is a later deliverable and will not impact the requirements gathering itself. The future of the core information profile, nonetheless, will depend on there at least being a discussion between funding bodies.

Risk: Wide differences in funder requirements

Likelihood (L): 2; **Impact (I)**: 1; **Factor (L X I)**: 2

Although there will be some diversity in funder requirements, it is not anticipated that the essential core of the required research information will differ much. By developing the core information profile in this project we can focus the efforts of other parts of UKRISS Phase 2 on the appropriate information.

Risk: Unable to clearly interpret funder requirements

Likelihood (L): 3; **Impact (I)**: 2; **Factor (L X I)**: 6

We anticipate that diverse funders will have unusual outlying requirements, which may be hard for the project to interpret. By defining the core information profile, we will focus our efforts on the most critical information, and avoid becoming hung-up on details that do not immediately benefit the work.

Risk: Funder requirements change over time

Likelihood (L): 5; **Impact (I)**: 2; **Factor (L X I)**: 10

In our Core work for this project, we will be aiming to capture the current funder requirements and to model them, on the assumption that the requirements do not change very much very fast. There is a risk that this assumption is wrong, but either way we will work initially with a “snapshot” of the funder requirements, with the view that the modelling process will be constructed to be accommodating to change in the long-run. In the more Advanced part of this work, we will further mitigate this risk by incorporating specific investigation into the changing requirements, and how those changes can be managed over time as the funders hopefully move towards harmonisation.

6.1.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core/Advanced): Requirements gathering

Timeline: Month 1 - Month 3 (or Month 6 if Advanced)
Owner: Kings College London, *Participants*: UKRISS project partners
Approach the funders and acquire their requirements and wish lists (if necessary).

WP2 (Core/Advanced): De-duplicate and document requirements

Timeline: Month 2 - Month 4 (or Month 7 if Advanced)
Owner: Kings College London, *Participants*: UKRISS project partners
Take the funders requirements and synthesise them into a consistent and documented information set (including all used schemas).

WP3 (Core): Derive core information profile

Timeline: Month 2 - Month 4
Owner: Kings College London, *Participants*: UKRISS project partners
From the full information set, extract the core information profile which provides the most common and regularly used fields in the sector

WP4 (Advanced): Gap analysis

Timeline: Month 5 - Month 7
Owner: Kings College London, *Participants*: UKRISS project partners
Carry out a gap analysis between the full information set and what can be actually provided by the UKRISS project partner institutions. This will also incorporate work to feed back the project findings to drive discussions around harmonisation.

Month	1	2	3	4	5	6	7
WP1							
WP2							
WP3							
WP4							

6.1.6 Existing work to consider

The key projects we will need to ensure we look at in order to ensure that we are being consistent with our information modelling include ROS, Research Fish, Gateway to Research and especially CASRAI.

6.2 Component 2: Serialising the core information profile

There is a lack of consensus/common agreement in the sector as to which the most critical research information is, and how it should be represented. The leads to difficulties in communication between similar systems and also difficulties in translating the data for re-use in other contexts (e.g. between VIVO-compliant and CERIF-compliant systems). The complexity of research information also presents a high barrier of entry for smaller organisations.

We will create serialisations of the core information profile developed under UKRISS Phase 2 using CERIF (and potentially VIVO), in order to improve the technical capacity for this information to be moved between systems.

The key benefit here will to be to provide a clear CERIF serialisation of the most important research information for use in related software systems and APIs, and to help drive harmonisation.

6.2.1 Drivers for this component

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community through harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D004 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

D006 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.2.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- Encourage a basic common set of standards for reporting, adopting existing ones where possible
- Drive alignment across the funders around a core dataset required to describe research
- Harmonise captured data for re-use in other contexts
- Minimise the effort/lower the entry barrier to comply with research information reporting
- Enable data exchange between systems, including providing crosswalks to other standards where possible/relevant (e.g. VIVO).

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.2.3 Approach and methodology

This section defines 2 pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra that could be carried out in addition to the Core work.

Core Work

Taking the core information profile produced in UKRISS Phase 2 we will carry out two key pieces of work to translate this into a CERIF model:

1. Determine the appropriate CERIF fields for the information and the way that those fields should be related together

2. Adopt where possible taxonomies for describing the relationships between the fields using the CERIF semantic layer, CASRAI and any other appropriate ontologies.

The result of this will be a profile of CERIF which details how to structure the Core Information Profile, which will be put forward to the community as a recommended way forward.

Advanced Work

Once the core information profile has been modelled to the degree possible, we will carry out a gap analysis between available ontologies and information/relationships which need to be described. It may be that some relationships cannot be described appropriately either with the CERIF standard as it is, or with the existing ontologies. In those cases we will make recommendations as to how to fix those issues, involving euroCRIS and other stakeholders in the process.

We will also build a prototype second serialisation of the information using the VIVO data model to demonstrate that it can be represented in other forms, and to indicate what the cross-walk between the CERIF and VIVO serialisations might be. VIVO has been selected because it is the only other significant emerging standard in the research information space, and has implications on how UK institutions might, in the future, exchange information with US organisations.

6.2.4 Feasibility and risks

Risk: Dependency on the core information profile

Likelihood (L): 3; **Impact (I)**: 5; **Factor (L X I)**: 15

Without the core information profile, the central premise of this project would not be possible. There is no clear mitigation for this, but in the event that it occurs, the project will divert its focus onto the existing taxonomies and standard definitions of important entities which are likely to be in the core information profile, in order at least to make some progress towards this goal.

Risk: Difficult/impossible to provide a clear CERIF model for the core information profile

Likelihood (L): 4; **Impact (I)**: 3; **Factor (L X I)**: 12

Given the flexibility of CERIF it is likely that there will be multiple ways to model the data, and multiple ontologies which could be useful to describe it. We will mitigate this risk in two ways: first, euroCRIS will be central to this project, as a source of definitive guidance on approaches to the modelling questions; second, we will consult all the relevant projects who have carried out modelling of CERIF entities and where possible simply adopt their approaches, to ensure consistency going forward.

Risk: Difficult/impossible to provide a clear VIVO model for the core information profile

Likelihood (L): 4; **Impact (I)**: 3; **Factor (L X I)**: 12

Given that the team is less familiar with VIVO, there is a risk that it does not do exactly what we need it to do, or that it is prohibitively hard to produce a model within the project timeline, or that it has too great a flexibility to point to a single approach. This problem will not be significantly mitigated, but in the event that this becomes an issue, the team will focus its effort on understanding VIVO more and documenting its experiences, with a view to potentially informing any further work.

6.2.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core): CERIF modelling and ontologies

Timeline: Month 1 - Month 4
Owner: Kings College London, *Participants*: euroCRIS
Take the core information profile and produce a CERIF model and set of ontologies to describe it.

WP2 (Advanced): Gap Analysis

Timeline: Month 4 - Month 5
Owner: Kings College London, *Participants*: euroCRIS
Perform a gap analysis between the core information profile and the CERIF model created in WP1

WP3 (Advanced): Define extensions to CERIF and ontologies

Timeline: Month 5 - Month 7
Owner: Kings College London, *Participants*: euroCRIS
Taking the gap analysis from WP2, define any required extensions to the CERIF model and any further ontologies required to describe the model, and make recommendations to the community.

WP4 (Advanced): VIVO modelling

Timeline: Month 1 - Month 4
Owner: Kings College London, *Participants*: euroCRIS
Take the core information profile and produce a VIVO model which describes it

Month	1	2	3	4	5	6	7
WP1							
WP2							
WP3							
WP4							

6.2.6 Existing work to consider

We will look at all projects in this sector which have been creating CERIF models/ontologies, including Gateway to Research, CASRAI, MICE, BRUCE, the RiO Extension project[31], and many others. We will also particularly look to extend the CERIF semantic model with the CASRAI terms/profiles where relevant.

6.3 Component 3: Enhanced intra-institutional, cross-institutional and institution-to-funder information exchange

Currently research information is managed with a mixed set of formats, with ad hoc reporting processes, which require a lot of manual intervention. There is a lack of common reporting processes and standards across all funders. This results in duplication of effort by all reporting institutions and presents a high entry barrier for smaller institutions.

We will use the system integration principles defined by the RMAS and RMAS tools to build a framework to combine and share information from a variety of institutional systems in appropriate formats (i.e. CERIF) to other systems (particularly funders). The RMAS infrastructure provides many tools to join systems together by providing flexible connectors that can connect systems within an institution together,

[31] RIO Extension: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/repositories/rioextension.aspx

and also provide means to connect to systems hosted externally. The external systems may include an Enterprise Service Bus (ESB) in the cloud, for example. Organisations can then easily connect and aggregate their systems, and make the data available to other systems both internally and externally.

The non-technical aspects of this project then centre around the policy aspects of the information flow and data exchange. There will be issues of trust, particularly if data goes outside the bounds of the institution, as well as issues of quality control of the data. We will need to explore what agreements might be required between institutions, and what internal organisational issues exist as barriers to this kind of approach.

The key benefit here is in streamlining the institution-to-funder reporting process, building on the harmonisation of the reportable information.

6.3.1 Drivers for this component

D001 Improve business intelligence, management and due diligence through better information quality and reporting utility

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community through harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D004 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

D006 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.3.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- A common set of reporting standards and processes across many funders/governing bodies is desired (RCUK, HESA, charities, and other funding bodies)
- A common set of research outputs and descriptions of those outputs is desired
- Smaller organisations with less in-house capacity for managing research information should be supported in their reporting processes
- There is a need to reduce the effort required to obtain and manage research information internally, and to increase options to validate captured information and to increase the automation in reporting this information to the funders (including for the REF)
- Internally institutions want to be able to combine data from HR, finance, student systems, CRISes and repositories
- A desire to reduce the number of systems to interact with, and the notion that there should be a “central” place through which research information reporting flows.

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.3.3 Approach and methodology

This section defines 2 pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra which could be carried out in addition to the Core work.

The overall approach of this project will be to construct a theoretical research project and study how the research information flows around it, and what policy issues are encountered along the way. We will, therefore, create example/test research information for the entities involved, such as: a project plan, research data, journal publications, sources of funding, funding agreements and terms, institutional affiliations, researchers, funder reporting requirements, institutional reporting requirements, etc. We will then, using the models developed elsewhere in UKRISS Phase 2 (where appropriate), prototype an environment which utilises the RMAS infrastructure to enhance the quality of the information flow, with a view to demonstrating the utility of that infrastructure in inter- and cross-institutional communication, and for reporting to funders.

We will then validate the desired information workflows by exploring the issues around sharing the data. There may be data protection issues around personal data, or concerns within institutions that they are exposing themselves to certain kinds of risk by sharing data. We will engage with key stakeholders throughout the project to ensure that these data governance and protection issues are addressed, and that proposed solutions are viable.

Core Work

The first part of the project will be to specify a suitably complex but sufficiently simple research project example, which will assume a project which is a collaboration between 2 institutions and is joint funded by 2 funding bodies. This will allow us to explore the interesting complexities which arise when the various reporting requirements for the different participants need to be met, within the framework of what is permissible by an institution.

At the same time, a UKRISS partner institution will deploy the RMAS infrastructure in a testing environment (and document their experiences), and be ready to implement the example project with that environment.

In the initial case, we will examine only a very small sub-set of the real research information, to make the problem tractable, and to focus on the information flow rather than the information itself. Later in the project we will look at the outputs of the core information profile work, and implement this within the RMAS set up at the partner institution, to demonstrate a more complete process.

The ultimate objective is to be able to submit to the 2 funders from the 2 institutions using this configuration; whether this is separate or combined reporting will be determined throughout the project. Our primary deliverables here are a detailed understanding of the information flow (which systems are involved, how they map to the core information profile, what the critical points and interactions are, etc), an implementation of this information flow using the RMAS infrastructure, and an understand of the barriers to this information flow caused by policy/legal concerns. By dividing the deliverables as such, it can be envisaged that the information flow can be realised in a more basic controlled environment without

the use of an institutional or cloud based ESB.

Advanced Work

A key advancement that can be made over the basic proof-of-concept information flow is to combine RMAS infrastructure and the Nexus cloud. We will work with Nexus[32] and JISC Advance to deploy a demonstration infrastructure with more participants (all of the UKRISS project partners, if possible), and show that by using the cloud we can further improve the information flow, by allowing an unbound number of participants either supplying or consuming the data. Nexus already have extensive experience here and can provide not only web services but also dashboards for providing access to the data in flexible ways, and may also be able to enhance the capacity to translate the information set into different schemas.

By this stage of the project we will have a clearer idea of the institutional policy issues which might prevent data going out to the cloud and we will therefore work within these bounds. At the same time we will attempt to identify ways in which data protection and competition concerns can be allayed, and determine if there is any technical infrastructure (e.g. security, standards compliance) which the cloud service provider could implement which would alleviate the issues.

As a final part to this project we will consider the feasibility of constructing a canonical CERIF database of research information in the UK based on information which passes through the cloud-based ESB. This is potentially a very large dataset, and potentially very difficult to achieve, with many complex challenges. It would be of value because canonical records for research information object (such as journals, publishers, etc) would be useful to normalise the data held in research information datasets around the country. We will explore the potential technical and policy challenges required to construct such a dataset and report on it.

6.3.4 Feasibility and risks

The main barrier is around the perception within organisations that they have more control over what is in/ out of their control with manual processes, and that changing that could expose them to risk.

Risk: Capacity to deploy infrastructure within partner institution's information services division

Likelihood (L): 3; **Impact (I)**: 2; **Factor (L X I)**: 6

Given that partner institutions will be bringing new work to the information systems division, there is a risk that their workload will impact the timeline of this project. This is mitigated by the fact that Brunel University has already begun to deploy the RMAS infrastructure, so are ahead of the curve on this project. Furthermore, we can imitate the integration with key business systems (e.g. HR) for the purposes of this proof-of-concept, and rely on running the RMAS infrastructure in a test environment during the project

Risk: Capacity to deploy infrastructure within the Nexus cloud

Likelihood (L): 3; **Impact (I)**: 2; **Factor (L X I)**: 6

We will be bringing work to Nexus, which has so far been outside the project, and therefore may have competing requirements and heavier workload than can accommodate the project. For this reason we have put this into the "Advanced" portion of this work, and will therefore have a longer lead time to do this, and it will not critically affect the rest of the project if it fails.

[32] JISC Advance Nexus: http://www.nexus.ac.uk/

Risk: Dependency on the core information profile

Likelihood (L): 3; **Impact (I)**: 2; **Factor (L X I)**: 6

Part of this project involves using the core information profile developed in UKRISS Phase 2 to show a more complete information flow. If delivery of this core information profile is delayed, this may delay this project. This is mitigated because we have separated the information from the information flow in this project, and therefore late delivery of the related project does not critically affect our work here.

Risk: Internal barriers for use of real institutional data within a cloud environment

Likelihood (L): 5; **Impact (I)**: 1; **Factor (L X I)**: 5

It is highly likely that institutions will not be allowed to share some or all of their data through the cloud, particularly in a proof-of-concept service. For that reason we plan to use test data in the main, potentially generated using software previously developed by the BRUCE project. This risk is not related to the long-term possibility that institutions will be unwilling to share their data in a full production requirement (see *Lack of wider take-up of RMAS*), only to the lifecycle of the UKRISS Phase 2 project.

Risk: Lack of wider take-up of RMAS

Likelihood (L): 3; **Impact (I)**: 3; **Factor (L X I)**: 9

If the sector as a whole does not adopt RMAS or an RMAS-like approach, then there is a risk that the work done here comes to nothing. The mitigation for this is that in this component we will study the information flow and processes required to significantly enhance reporting to funders, and that these processes could be re-deployed within an alternate environment if RMAS is non-viable in the long run (i.e. the flow of information is a key output of this component). Furthermore, developing a demonstrator using RMAS may improve its chances of being more widely adopted in future.

6.3.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core): Institutional deployment of RMAS infrastructure

Timeline: Month 1 - Month 3

Owner: Brunel University, *Participants*: University of Exeter

Brunel University will deploy and configure to the greatest extent possible the RMAS infrastructure in preparation for working with the project.

WP2 (Core): Preparing the information flow demonstrator

Timeline: Month 1 - Month 4

Owner: University of Exeter, *Participants*: UKRISS project partners

We will construct the test project for use in the demonstrator and build all the relevant mappings for RMAS to support the required information.

WP3 (Advanced): Integration with Nexus cloud

Timeline: Month 1 - Month 7

Owner: University of Exeter

From the very start of the project interact with Nexus and JISC Advance to prepare for and then to deploy the RMAS infrastructure for the demonstrator on their cloud.

WP4 (Advanced): Feasibility study for CERIF-in-the-sky

Timeline: Month 5 - Month 6 (or any other 2 month period within the project)

Owner: University of Exeter, *Participants*: UKRISS project partners

Carry out a feasibility study of building a large CERIF database during normal working procedures of a cloud-based research information reporting system

Month	1	2	3	4	5	6	7
WP1							
WP2							
WP3							
WP4							

6.3.6 Existing work to consider

The main body of this project will be based on the RMAS infrastructure. In terms of information modelling, there are many other efforts which we may wish to look at, including IRIOS[33]/IRIOS2[34], CERIF In Action[35], CERIF4REF[36], EXRI-UK, etc. although we will not go far down the information modelling route in this project, focussing instead on the information flow.

6.4 Component 4: Shared reporting/benchmarking

The key use for research information within an institution is for management information, for which there is limited tooling, standards or good data. This is currently handled by manual intervention. There are management information benefits to be had from improving benchmarking against other institutions, or within institutions against Key Performance Indicators (KPIs). There is also an opportunity to reduce the overall burden on research information managers.

This project therefore proposes to develop software (started under the BRUCE) to consume CERIF and provide a reporting interface, and to add a comparative/benchmarking component to it. This will be informed by some requirements gathering and consultation with the Snowball Project which has analysed the particular issues around data governance and sharing for benchmarking. Stakeholders will be consulted throughout the project to ensure that their needs/concerns are taken into account.

The key benefit here is to improve management information and business intelligence through software and processes within an institution.

6.4.1 Drivers for this component

D001 Improve business intelligence, management and due diligence through better information quality and reporting utility

D002 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D003 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

[33] IRIOS: http://www.irios.sunderland.ac.uk/

[34] IRIOS2: http://irios2.wordpress.com/

[35] CERIF in Action (CiA): http://cerifinaction.wordpress.com/

[36] CERIF4REF: http://r4r.cerch.kcl.ac.uk/?p=113

D004 Improve the research, strategy and planning across UK institutions through use of better quality reporting information

D005 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.4.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- Provide benchmarking in a variety of contexts: between institutions based on data from funders, within institutions (e.g. department to department, group to group, or any other sub-sections of the university) and against KPIs
- Provide a reporting interface which brings in data from a variety of institutional systems such as HR, finance, CRISes, the repository, etc.
- Make the research metrics of the reports available in a standard and consistent way
- Provide a variety of different views on the research information (and allow combinations of views): researcher-centric, institution-centric, subject-centric, department-centric, cross-institutional, etc.
- Support a wide variety of research outputs, areas of interest, and support understanding collaborations between researchers
- Be flexible enough to offer ad-hoc report generation which may be produced on-the-fly or periodically.

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.4.3 Approach and Methodology

This section defines 2 pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra could be carried out in addition to the Core work.

The project will adopt for its basis the outputs of the BRUCE project, which was a JISC-funded initiative to prototype a dynamic reporting/management information interface on top of a search index which was seeded with data derived from a CERIF dataset. The project produced two key outputs that we are interested in here:

1. The SolrEyes[37] user interface, a faceted browse application set up to allow the user to explore a research information dataset and to produce tabular reports which fulfilled the internal reporting requirements of Brunel University.
2. A methodology and a model for taking a highly relational CERIF dataset and converting it into a format which can be indexed and reported upon by SolrEyes.

[37] SolrEyes (also known as Edjo): https://github.com/CottageLabs/edjo

This project proposes to take these prototypes and extend them to enable cross-institutional reporting, and intra- and cross- institutional benchmarking within the feature set of the software, and to demonstrate the value-add that this can provide for management information provision.

The resulting prototype software will be usable within an institution for their own information, or within a funding body for providing a view across their own dataset. It could also be run as a service on behalf of one or more institutions, or the sector as a whole.

In parallel to this project, there is a JISC-funded Gateway to Research[38] initiative which is using the outputs of the BRUCE project to index and present data from GtR's API. This is highly complementary to this proposal, and we will be able to share reporting requirements and code with that project. Our focus here will be on the additional data beyond that which comes from GtR and incorporating that into management information reports, but being able to incorporate GtR data as well will further enhance the system. Our focus is also on benchmarking and comparative information, which may be of long-term value to the GtR project also.

Core Work

In order to be valuable, the system needs to provide views over the dataset from a number of perspectives: the researcher, their department, their institution, a funding body, etc. We will, therefore, initially identify the key perspectives from the UKRISS requirements analysis, and develop mappings from CERIF to an indexable format which provides the appropriate view on the data.

We will also, at this point, carry out a study within the UKRISS partner institutions into the data they would be willing to share for benchmarking against other institutions. This should allow us to define a set of requirements for our index which will be of a suitable granularity; for example, it is unlikely that institutions would be willing to share researcher-level research information with other institutions for benchmarking purposes, as this might damage their competitiveness. We will try to extract such requirements and build a benchmarking index which is sensitive to the balance between the desire to do accurate benchmarking and the desire to maintain competitive advantage.

Where appropriate, this work will draw on the other projects in UKRISS phase 2 (for example the CERIF modelling and minimum/maximum information set work) to provide the model which will be indexed, as well as reporting requirements coming out of Project Snowball[39], Engage[40] and JISC-GtR. We will also speak to representatives of the research office at each UKRISS partner institution, in order to validate the reports that we plan to generate.

As part of this process we will develop software which can take an input CERIF dataset, apply the transformation to an indexable form, and build the resulting index. Where appropriate this will be supported by a suitable administrative user interface, which will then provide the back-end part of the SolrEyes software.

Once we have suitable mechanisms for converting CERIF data into an indexable and reportable form, we will build two essential front-end components:

---

[38] G4HE: there is no reference for this at time of writing, but see http://www.jisc.ac.uk/blog/reporting/

[39] Snowball: http://www.projectsnowball.info/

[40] Engage: http://researchclusters.wordpress.com/

1. More tabular reports similar to those created in the BRUCE project, representing the different perspectives that the data will be viewed from (e.g. a researcher-centric report, a department-centric report, etc)
2. A comparative tabular interface which will allow the selection of two or more reporting units (e.g. departments) and their resulting reports to be presented together in an appropriate interface for comparison.

By this stage in the project, then, we will have a consistent way of generating an index and the basis of a powerful way of interacting with the data in the index for management information purposes.

Given that obtaining the data for indexing is not strictly part of this project, we will work - where necessary - with generated test CERIF data (as demonstrated possible by the BRUCE project) which meets the requirements of the other models developed by UKRISS Phase 2.

Advanced Work

There are two main areas of work which would significantly enhance the management information software produced by this project:

1. A mechanism for continually keeping the data in the index up to date
2. More advanced graphical representations of the reporting data, such as comparative, interactive line charts, etc

The advanced part of this project will therefore look for a mechanism by which research information from a variety of systems can be integrated into the index, and implement a continuous pipeline of data into the system.

This will most likely draw from the sister UKRISS Phase 2 project, which is using the RMAS infrastructure to combine up research information to make it available for institutions and funders. If component is taken to its **Advanced** stage, there will also potentially be benefit which can be drawn from tools provided in the Nexus cloud.

Meanwhile, the graphical user interface will be developed by adopting a graphic charting software library or service (such as Google Charts[41]) and implementing the existing tabular reports and comparisons as interactive graphs.

We will also look for desirable user interface and user experience enhancements by gathering feedback from potential users of the system at Brunel University and other UKRISS partner institutions. This will be via contacts with the research offices at each institution, and we will be aiming to have as many senior management users as possible give their view on the system, since they will most often be the “customer” for the reports it generates.

6.4.4 Feasibility and risks

Overall, the software development portions of this project have a high feasibility and a low risk, as they are based on principles and code developed in previous projects. The main source of risk is in the ability to obtain good data for the system; this is a general problem. The project deals with it by isolating the software development from the data through the use of a test dataset, and defers to other UKRISS

[41] Google Charts: https://developers.google.com/chart/

Phase 2 components to enhance the availability and flow of data. At the same time we are keeping the route open for manual acquisition of data as is currently done when generating management information reports.

Risk: Unable to obtain all the relevant data for reporting

Likelihood (L): 5; **Impact (I)**: 3; **Factor (L X I)**: 15

This is our key risk, and is quite likely since good processes for sharing this information do not yet exist. We will mitigate this within the project by first using a test dataset so that software development can proceed without being affected. Once the software is developed it is hoped that combined with other UKRISS Phase 2 work and the demonstrable utility of the management information that the relevant data for reporting will become easier to obtain.

Risk: Reporting requirements are too diverse

Likelihood (L): 2; **Impact (I)**: 2; **Factor (L X I)**: 4

The project will focus its efforts on the reports generally required for management information. Having already looked at this in the BRUCE project, we know that the reporting requirements do not differ significantly across institutions. Furthermore, the objective of the SolrEyes interface is to provide a flexible way of building the desired reports, so some diversity is comfortably tolerated.

Risk: CERIF data models are not stable enough to index

Likelihood (L): 4; **Impact (I)**: 3; **Factor (L X I)**: 12

A fully stable outcome of this project depends on the ability to map from CERIF to an indexable/ reportable form. If the CERIF model is too loose or insufficiently well defined it may be difficult to stabilise that mapping. We will mitigate this by ensuring that the mapping is highly configurable, so that it can easily incorporate changes from other UKRISS Phase 2 projects as they become available, or so that institutions can customise the software to their own needs.

Risk: No method of continually integrating new data can be found

Likelihood (L): 4; **Impact (I)**: 3; **Factor (L X I)**: 12

With many different systems involved, there is a high likelihood that the automatic acquisition of data to be integrated into the index will not run smoothly. To mitigate this we will ensure (as was done in BRUCE) that the process of obtaining and indexing data manually is documented and supported, so that even if not all systems can be integrated, there is a route to including the data manually at equivalent or lower cost than existing manual reporting processes.

6.4.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core): Reporting requirements and CERIF model

Timeline: Month 1 - Month 2
Owner: Brunel University, *Participants*: UKRISS partners

Define the different report types that are required (e.g. researcher-centric, funder-centric), the requirements/limitations of benchmarking use cases, and determine the fields/relationships in CERIF from which the data for these reports can be found.

WP2 (Core): CERIF to index mapping

Timeline: Month 2 - Month 3
Owner: Cottage Labs, *Participants*: Brunel University
Take the desired reports as defined in WP1 and define a mapping to an indexable/reportable form for use by SolrEyes. This will involve writing the software to make this translation happen and be re-runnable.

WP3 (Core): Tabular reporting/benchmarking

Timeline: Month 3 - Month 7
Owner: Cottage Labs
Develop the user interface features required to present tabular reports and comparative benchmarking tables.

WP4 (Advanced): Continuous data integration

Timeline: Month 3 - Month 7
Owner: Cottage Labs, *Participants*: Brunel University, University of Exeter
Develop a mechanism based on the sister UKRISS Phase 2 project to continually update the SolrEyes index so that management information is as current as possible.

WP5 (Advanced): Graphical reporting/benchmarking

Timeline: Month 5 - Month 7
Owner: Cottage Labs
Develop the user interface features required to present graphical reports and comparative benchmarking.

Month	1	2	3	4	5	6	7
WP1							
WP2							
WP3							
WP4							
WP5							

6.4.6 Existing work to consider

This project will be based heavily on the BRUCE project for the **Core Work**, as well as reporting information output from Project Snowball, Engage and JISC-GtR.

In the **Advanced Work**, the developments of the RMAS project will be highly valuable.

6.5. Component 5: Standardising research information terms

There is a lack of consistency and understanding across the sector of some “common” terms in use in research information (for example, expressions of metrics, researcher roles, research entities, etc). This leads to difficulties in clear communication between systems, even if they are standards compliant.

This would, therefore, be a consensus building exercise where the various terms and dictionaries are reviewed and a common way forward is determined. This is potentially a large amount of work, so the project would focus on the most essential areas (as defined by the core information profile) and the output would be a strong case for the community to follow the recommendations. In order to achieve this, we will need to engage a broad spectrum of stakeholders to validate the recommendations, and to provide feedback throughout the work.

The key benefit here is as an enabler for other work around research information, and the harmonisation of terms across the sector.

6.5.1 Drivers for this component

D001 Improve business intelligence, management and due diligence through better information quality and reporting utility

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community through harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D004 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

D006 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.5.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- Capture and present research metrics in a consistent way which is suitable for benchmarking
- Drive a shared agreement across all funders for the meaning of particular terms
- Define or recommend taxonomies for researcher roles, research entities, domains/themes, institutional structures, output types, etc.
- Define or recommend standard forms of certain entities, such as dates, funding references (e.g. in support of FundRef), etc.
- Drive the adoption of existing standards (e.g. MeSH).

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.5.3 Approach and methodology

This section defines 2 pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra could be carried out in addition to the Core work.

Core Work

Taking only the **core information profile** produced in UKRISS Phase 2 we will identify key standard components which would benefit from being represented consistently across the sector (e.g. date formats, names of research outputs, etc), and carry out a consensus-building exercise to recommend some standards or identify where there is consensus (i.e. it is not anticipated that there will be much new work in this component). This will involve looking at all projects which are working (or have been working) in this sector which have produced profiles or other recommendations which are relevant (see **Section 4.7** for a summary), in particular the work of the Snowball project and CASRAI. We will also contact a broad spectrum of stakeholders for feedback/input to the consensus-building exercise, both to inform the work, and to improve its likelihood of long-term adoption.

The output of this project will be a documented set of recommendations to the community for discussion and potential adoption.

Advanced Work

There is no advanced work in this component.

6.5.4 Feasibility and risks

Risk: Dependency on the core information profile

Likelihood (L): 3; **Impact (I)**: 5; **Factor (L X I)**: 15

Without the core information profile, the central premise of this project would not be possible. There is no clear mitigation for this, but in the event that it occurs, the project will divert its focus onto the CERIF standard itself and identify important entities which are likely to be in the core information profile, and carry out the consensus building around them, in order at least to make some progress towards this goal.

Risk: Scope of work is too broad

Likelihood (L): 4; **Impact (I)**: 1; **Factor (L X I)**: 4

Given the large amount of information in this sector, there is a high likelihood that the scope of work will be very broad. We will mitigate this by focussing the project on only those terms which come out of the core information profile; this will ensure that we not only limit our scope, but also address the issues most pressing to the community.

6.5.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core): Identify key fields/information

Timeline: Month 1
Owner: Kings College London, *Participants*: euroCRIS
Take the core information profile and identify the entities within which could benefit from standardisation

WP2 (Core): Standardisation

Timeline: Month 2 - Month 7
Owner: Kings College London, *Participants*: euroCRIS

Carry out the consensus building exercise for as many of the fields/information points as possible. This work package could expand or contract to consume all available time, depending on the level of detail required.

Month	1	2	3	4	5	6	7
WP1							
WP2							

6.5.6 Existing work to consider

We will look at all projects in this sector which have been creating CERIF models/ontologies, including Gateway to Research, CASRAI, MICE, BRUCE, the RiO Extension project, and many others.

6.6 Component 6: Institutional data CERIF model

The flexibility of CERIF means that without further profiling information it is still non-trivial for institutions/ funders to exchange research information reliably and consistently regarding researchers, departments, cost centres, etc (and to uniquely identify them). It is also currently quite difficult to combine HR data with other kinds of data (requiring a lot of manual intervention), because of the varying expressions of the data.

We will, therefore, define or adopt a CERIF model for representing data important to an institution's understanding of its research information (in particular HR and institutional structure). There is a need to be able to track researchers within their own institution and between organisations (and not just universities, but also as they move out of academia and into the broader public sector and the private sector).

The key benefit here is as an enabler for other work around research information, and in driving harmonisation across the sector.

6.6.1 Drivers for this component

D001 Improve business intelligence, management and due diligence through better information quality and reporting utility

D002 Reduce the reporting burden and increase the efficiency or response agility of the research community through harmonisation of reporting processes and/or systems

D003 Enable cross-sector impact analysis, evaluation and strategy development through systemisation and harmonisation of reporting

D004 Increase research community reporting compliance through deploying easy-to-use, flexible reporting systems with user benefits

D006 Improve research information management across the sector through deploying sustainable, affordable solutions that are fit-for-purpose

6.6.2 Requirements to be addressed by this component

The community requirements behind this work broadly fall into the following high-level descriptions:

- Be able to provide a researcher-centric and group-centric view of the research information within an institution
- Be able to combine HR data easily with data from other systems (such as Finance, the CRIS, repositories, etc)
- Provide unique identifiers for entities such as people, organisational units and cost centre
- Provide a model for describing organisational structures, and support describing their change over time
- Support detailed relationship analysis between researchers and organisations
- Comply with HESA requirements.

Note that this component works towards meeting these requirements, and will not necessarily meet any of them fully. See **Appendix H** for a complete list of requirements, and **Appendix J** for a full treatment of relationships between components and requirements/drivers.

6.6.3 Approach and methodology

This section defines 2 pieces of work which could be carried out by the project. The **Core Work** is the essential part of the project, and the **Advanced Work** is an optional extra could be carried out in addition to the Core work.

Core Work

Some of the required modelling work has already been done or at least begun by existing projects (see below), so we plan to bootstrap this work by adopting approaches from them (e.g. RMAS, CERIF in Action, etc). In particular they have done a lot of prior work on modelling HR feeds, so we will pull that out of that project and incorporate it into the modelling of both HR feeds and institutional structures. At this stage we will test and validate our work with the UKRISS project partners only.

We will then be able to carry out a gap analysis between the data that we can currently model, and the actual data within institutions and as required by funders. We will export HR data from the UKRISS partner institutions in the formats defined by this project and compare it with the full Information Profile determined in UKRISS Phase 2, to understand where further modelling is required. We will also incorporate into this process the HESA requirements for reporting on organisational structures.

Advanced Work

Where the gap analysis presents us with further modelling work we will work with euroCRIS to define a good model for representing this institutional data and propose that to the community. We will also engage further stakeholders to validate the model and gap analysis with them, and also improve chances of long-term adoption.

We will also look further at the information available within institutions outside of the core HR fields and those required by the full information profile, to see if there is additional value to be had by incorporating it into the model. If so we will attempt to expand the model further, bearing in mind that the number of fields in an HR database could be very large, and that this could therefore be a lot of work.

6.6.4 Feasibility and risks

Risk: Definition of the full Information Profile

Likelihood (L): 3; **Impact (I)**: 3; **Factor (L X I)**: 9

This project partially depends on the development of the full information profile, so if it is delayed in its delivery that will have a knock-on effect. We can mitigate that by focussing our work on the data that the institution has and wants to use internally, which is conceptually independent from their reporting requirements; the reporting requirements in the form of the information profile mostly provide validation that we are focussed on the most useful data within the institution, so we will simply need to ensure that we limit ourselves to working on fields which multiple UKRISS partners agree are important. The gap analysis, nonetheless, will be impacted.

Risk: Timescale for usable output

Likelihood (L): 3; **Impact (I)**: 2; **Factor (L X I)**: 6

Because of the potentially large number of fields, it is possible that scope creep may affect the output of this project. By focussing on the reporting requirements of the Information Profile and by validating that a particular field is important to more than one UKRISS partner we can control that scope creep, and ensure that the deliverables are on time.

Risk: Breadth of study is not sufficiently wide

Likelihood (L): 4; **Impact (I)**: 2; **Factor (L X I)**: 8

Because we will be basing our work only on the data held by the UKRISS partners, there is a risk that we will not properly understand all the kinds of internal data in use across the sector within different types of institution. This will not be mitigated, but will be accepted as a necessary limitation of the project, and documented as such. Final outputs will therefore need to be validated by further work before being applicable to the whole community.

6.6.5 Workpackages and timeline

All recommendation projects will be managed under UKRISS WP1, so this section just defines packages of tasks with timelines, appropriate milestones and task owners. Workpackages are also identified as belonging to the **Core Work** or to the **Advanced Work**.

WP1 (Core): Obtain requirements from the Information Profile and HESA

Timeline: Month 1 - Month 3
Owner: Brunel University, *Participants*: UKRISS project partners
Obtain from the Information Profile developed in UKRISS Phase 2 and from HESA directly the requirements needing to be met by an institutional data model

WP2 (Core): Obtain example data from UKRISS partners

Timeline: Month 1 - Month 3
Owner: Brunel University, *Participants*: UKRISS project partners
Obtain from each UKRISS project partner institution some example HR and organisational data, and use cases for how this data would ideally be used

WP3 (Core): Gap analysis on institutional data

Timeline: Month 3 - Month 4
Owner: Brunel University, *Participants*: UKRISS project partners
Perform a gap analysis between the required data and the data available at institutions

WP4 (Advanced): CERIF modelling of gaps

Timeline: Month 4 - Month 7
Owner: Kings College London, *Participants*: euroCRIS
Work with euroCRIS to define ways of plugging the gaps in support for institutional data

WP5 (Advanced): Further institutional data requirements

Timeline: Month 1 - Month 3
Owner: Brunel University, *Participants*: UKRISS project partners
Identify more unusual institutional data and see if it can be incorporated into the CERIF model.

Month	1	2	3	4	5	6	7
WP1							
WP2							
WP3							
WP4							
WP5							

6.6.6 Existing work to consider

There is extensive prior work by RMAS and BRUCE, as well as it being important to look at the work of ORCID, ISNI[42] and CASRAI, and also at the behaviour and data of the main CRIS vendors' software. There are also a number of relevant JISC-funded projects such as Cerif In Action, Readiness4Ref, CRISPool[43], CERIFy[44], and IRIOS.

[42] ISNI: http://www.isni.org/
[43] CRISPool: http://www.crispool.org/
[44] CERIFy: http://cerify.ukoln.ac.uk/

7. Conclusions and recommendations

7.1 Project team conclusions and recommendations

As we have seen, there is a lot of effort going into the research information space from a variety of stakeholders, both public and private sector. We have spoken to many of these stakeholders and gained deep insights into what drives their business decisions and what their requirements are to meet those drivers.

The overall consensus is that the UK does not require a national research information system, and that such a system would be extremely hard to build and be adopted. Instead there is a focus on harmonisation of data formats, and the use of common APIs to allow systems to connect together in useful and meaningful ways.

The components that we detailed in **Section 6**, and those further listed in **Appendix J** represent the kinds of developments that are therefore demanded by the sector based on those requirements and drivers. From a list of over twenty possible developments, we chose the six presented in the main body of this report based on the perceived need for them and whether there was significant effort going into them at the moment (independently of the UKRISS project). Each of those selected is regarded as being both important and currently under-investigated.

From those components, we were able to construct three options (**Section 5**) which reflect feasible combinations of those components based on the three main identified work areas: information modelling, benchmarking and reporting infrastructure.

It therefore falls to the project to make a recommendation to JISC and the Steering Board as to the way forward into Phase 2 and beyond. We recommend **Option 3: Focus on Reporting**.

The reasons behind this recommendation are:

1. It has a compelling business case justifying immediate implementation, with a broad set of beneficiaries
2. It provides infrastructure which can be re-used in a variety of other useful contexts, such as in more advanced integration of benchmarking tools
3. It meets a range of requirements and sits comfortably with many of the drivers
4. It will, within the bounds of Phase 2, produce some working proof-of-concept code and systems which will be potentially viable to take forward into a production environment beyond the end of the project.

We believe that this is therefore the most appropriate way forward for the project.

7.2 Steering board comments and recommendations

The UKRISS Steering Board were presented with the final draft version on 30th November and met on 6th December to review the report and make a recommendation to JISC based on the three options presented in the report.

Specific comments made on each of the options were as follows.

Option 1 (Focus on modelling)

This option was seen as important and fundamental to simplifying research reporting and increasing the exchange of research information across the sector. It was also seen as an open-ended task, and that it would be difficult to achieve the necessary consensus on a core information profile within the limited timescales of the project. Putting more effort into aligning core information fields was seen to be more relevant to the project objectives than extending the CERIF standard into new areas. There was a clear distinction made between quantitative reporting, such as outputs, and qualitative reporting such as narrative text. The opportunity for developing the core profile was seen to be greatest around the quantitative reporting fields. The work on data dictionaries was also viewed as very relevant.

The output of this option would be a proposal for standardisation, which would be less tangible than a demonstrator. This would potentially make adoption of the outcomes more difficult. Work on the information profile should consider input from funders on fields that are considered essential in order to rationalise and simplify reporting requirements.

Option 2 (Focus on benchmarking)

Benchmarking was the least favoured option. There is considerable overlap with other initiatives such as the Snowball project. Providing tools to support benchmarking was seen as very useful by the Steering Board, which was also reflected by the large number of requirements collected in the study. Benchmarking was seen as an endpoint, with many other issues to be addressed before such as access to high quality input data. The benchmarking option was also seen as furthest from the original objectives of UKRISS.

Option 3 (Focus on reporting infrastructure)

Development of infrastructure to support reporting was seen as the preferred option by the Steering Board. Given the timescales of the project, the outcome would be a demonstrator and more detailed report, rather than a production system. There was a request to put less emphasis on the benchmarking component and direct more resources to the modelling and infrastructure tasks. There was seen to be a potential overlap with the JISC Gateway to Research project, which is considering the flow of reporting information from funders to institutions.

Engagement with the sector was seen as vital to gain buy-in for a national service. Further development of the business case would be required, particularly around the strategic value to the sector and the long-term benefits. The costs of such a service would need to be clearly defined, so it is clear that they will not provide a barrier to adoption. More ambitious goals for uptake of a national service would be required, which should be considered in the sustainability work in Phase 2.

Recommendations summary

There was strong consensus amongst the Steering Board that option 3 should be recommended to JISC as the preferred direction for Phase 2. Option 2 was the least favoured option, particularly due to the overlap with other initiatives and also the change of direction for UKRISS this would imply. Option 1 was seen as fundamental, and although contributions could be made, it would be difficult to achieve a tangible outcome within the timescales of the project. The international perspective was also seen as important and should be taken into account in all Phase 2 work.

8. Glossary

Table 8.1 contains a list of frequently used acronyms and terms followed by a brief description and link to further information.

Acronym or term	Description
API	Application Programming Interface http://en.wikipedia.org/wiki/Application_programming_interface
ARMA	Association of Research Managers and Administrators UK http://www.arma.ac.uk/
Atira	CRIS vendor http://www.atira.dk/
Avedas	CRIS vendor http://www.avedas.com/
BIS	Department of Business Innovation and Skills http://www.bis.gov.uk/
CASRAI	The Consortia Advancing Standards in Research Administration Information (CASRAI) is a non-profit standards development organisation http://casrai.org/
CERIF	Common European Research Information Format http://www.eurocris.org/cerif/introduction
Converis	Current Research Information System for Higher Education Institutions from Avedas http://www.avedas.com/en/converis.html
CRIS	Current Research Information System http://www.wikipedia.org/wiki/Current_Research_Information_System/ See also: http://www.cris2010.org/
CrossRef	Official DOI® link registration agency for scholarly and professional publications http://www.crossref.org/
DOI	Digital Object Identifier http://www.doi.org/
Dublin Core	Dublin Core Metadata Initiative http://www.dublincore.org
Elements	Publications management system from Symplectic http://www.symplectic.co.uk
ESB	Enterprise Service Bus: http://en.wikipedia.org/wiki/Enterprise_service_bus
EPSRC	Engineering and Physical Sciences Research Council http://www.epsrc.ac.uk/
euroCRIS	A not-for-profit association which aims to be the internationally recognised point of reference for all matters relating to Current Research Information Systems (CRIS) http://www.eurocris.org/
FOI	Freedom of Information http://en.wikipedia.org/wiki/Freedom_of_information_laws_by_country
FTE	Full-time equivalent

FundRef	Collaborative pilot project of scholarly publishers and funding agencies, facilitated by CrossRef, to provide a standard way of reporting funding sources for published scholarly research http://www.crossref.org/fundref/index.html
Gateway to Research (GtR)	Access to Research Council funded research information http://www.rcuk.ac.uk/research/Pages/gtr.aspx
Grants on the Web	Research Council grants information portals e.g. http://gow.epsrc.ac.uk/
HE-BCI	Higher Education Business and Community Interaction survey http://www.hefce.ac.uk/whatwedo/kes/measureke/hebci/
HEFCE	Higher Education Funding Council for England http://www.hefce.ac.uk/
HEI	Higher Education Institution
HESA	Higher Education Statistics Agency http://www.hesa.ac.uk
InCites	Thomson Reuters research analytics tool http://researchanalytics.thomsonreuters.com/incites/
ISNI	International Standard Name Identifier http://www.isni.org/
J-eS	Joint Electronic Submission system – for submission and monitoring of research grant applications to the Research Councils https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx
JISC	The funder of the UKRISS project http://www.jisc.ac.uk/
KPI	Key Performance Indicator
MeSH	Medical Subject Headings http://en.wikipedia.org/wiki/Medical_Subject_Headings
METIS	Dutch research information database system http://www.metis.hosting.kun.nl/metis/
MRC	Medical Research Council http://www.mrc.ac.uk/index.htm
NERC	Natural Environment Research Council http://www.nerc.ac.uk/
ORCID	Open Researcher Contributor Identification Initiative http://about.orcid.org/
Pure	CRIS software produced by Atira http://www.atira.dk/en/pure/
PVC	Pro-Vice Chancellor
RCUK	Research Councils UK http://www.rcuk.ac.uk/Pages/Home.aspx
REF	Research Excellence Framework http://www.ref.ac.uk/
Research Fish/e-Val	A research outcomes system designed to enable a researcher to report once across multiple funders https://www.researchfish.com/
RMAS	Research Management & Administration System (http://www.rmas.ac.uk/)
ROS	Research Outputs System http://www.rcuk.ac.uk/media/news/2011news/Pages/111124.aspx
SolrEyes	Faceted browse technology from the BRUCE project
STFC	Science & Technology Facilities Council http://www.stfc.ac.uk/
Symplectic	CRIS vendor http://www.symplectic.co.uk
UCISA	Universities and Colleges Information Systems Association http://www.ucisa.ac.uk/

VIVO	Research information portal http://vivoweb.org/
Wellcome Trust	Health charity funder http://www.wellcome.ac.uk/

Table 8.1: Glossary of terms